



International Journal of Advanced Computer Science and Applications

Volume 2 Issue 12

December 2011



ISSN 2156-5570(Online)

ISSN 2158-107X(Print)



www.ijacsa.thesai.org



W H E R E W I S D O M S H A R E S

INTERNATIONAL JOURNAL OF
ADVANCED COMPUTER SCIENCE AND APPLICATIONS



THE SCIENCE AND INFORMATION ORGANIZATION

www.thesai.org | info@thesai.org





INTERNATIONAL JOURNAL OF
ADVANCED COMPUTER SCIENCE AND APPLICATIONS



A Publication of
The Science and Information Organization



IJACSA Editorial

From the Desk of Managing Editor...

It is a pleasure to present our readers with the December 2011 Issue of International Journal of Advanced Computer Science and Applications (IJACSA).

The renaissance stimulated by the field of Computer Science is generating multiple formats and channels of communication and creativity. IJACSA is one of the most prominent publications in the field and engaging the ubiquitous spread of subject knowledge with effectiveness in all classes of audience. Nevertheless, the promise of increased engagement requires that we consider how this might be accomplished, delivering up-to-date and authoritative coverage of advanced computer science and applications.

The journal has a wide scope ranging from the many facets of methodological foundations to the details of technical issues and the aspects of industrial practice. It includes articles related to research findings, technical evaluations, and reviews. In addition it provides a forum for the exchange of information on all aspects.

The editorial board of the IJACSA consists of individuals who are committed to the search for high-quality research suitable for publication. These individuals, working with the editor to achieve IJACSA objectives, assess the quality, relevance, and readability of individual articles.

The contents include original research and innovative applications from all parts of the world. This interdisciplinary journal has brought together researchers from academia and industry as well as practitioners to share ideas, problems and solutions relating to computer science and application with its convergence strategies, and to disseminate the most innovative research. As a consequence only 29% of the received articles have been finally accepted for publication.

Therefore, IJACSA in general, could serve as a reliable resource for everybody loosely or tightly attached to this field of science.

The published papers are expected to present results of significant value to solve the various problems with application services and other problems which are within the scope of IJACSA. In addition, we expect they will trigger further related research and technological improvements relevant to our future lives.

We hope to continue exploring the always diverse and often astonishing fields in Advanced Computer Science and Applications.

Thank You for Sharing Wisdom!

Managing Editor

IJACSA

Volume 2 Issue 12, December 2011

editorijacsa@thesai.org

ISSN 2156-5570 (Online)

ISSN 2158-107X (Print)

©2011 The Science and Information (SAI) Organization

Editorial Board

Dr. Kohei Arai – Editor-in-Chief

Saga University

Domains of Research: Human-Computer Interaction, Networking, Information Retrievals, Optimization Theory, Modeling and Simulation, Satellite Remote Sensing, Computer Vision, Decision Making Methodology

Dr. Ka Lok Man

Xi'an Jiaotong-Liverpool University (XJTLU)

Domain of Research: Computer Science and Microelectronics

Dr. Sasan Adibi

Research In Motion (RIM)

Domain of Research: Security of wireless systems, Quality of Service

Dr. Zuqing Zuh

University of Science and Technology of China

Domains of Research : Optical Communication Systems, Optical network architecture and design, Next generation Internet, Signal processing, Broadband access network, such as cable access (DOCSIS) networks, passive optical networks (PON), fiber to the home (FTTH), Energy-efficient network and green technologies

Dr. Sikha Bagui

University of West Florida

Domain of Research: Database, database modeling, ER diagrams, XML data, web databases, data mining, association rule mining, data preprocessing

Dr. T. V. Prasad

Lingaya's University

Domain of Research: Bioinformatics, Natural Language Processing, Image Processing, Robotics, Knowledge Representation

Dr. Mohd Helmy Abd Wahab

Universiti Tun Hussein Onn Malaysia

Domain of Research: Data Mining, Database, Web-based Application, Mobile Computing

IJACSA Reviewer Board

- **A Kathirvel**
Karpaga Vinayaka College of Engineering and Technology, India
- **Abbas Karimi**
I.A.U_Arak Branch (Faculty Member) & Universiti Putra Malaysia
- **Dr. Abdul Wahid**
Gautam Buddha University, India
- **Abdul Khader Jilani Saudagar**
Al-Imam Muhammad Ibn Saud Islamic University
- **Abdur Rashid Khan**
Gomal University
- **Dr. Ahmed Nabih Zaki Rashed**
Menoufia University, Egypt
- **Ahmed Sabah AL-Jumaili**
Ahlia University
- **Md. Akbar Hossain**
Aalborg University, Denmark and AIT, Greeceas
- **Albert Alexander**
Kongu Engineering College,India
- **Prof. Alcinia Zita Sampaio**
Technical University of Lisbon
- **Amit Verma**
Rayat & Bahra Engineering College, India
- **Ammar Mohammed Ammar**
Department of Computer Science, University of Koblenz-Landau
- **Arash Habibi Lashakri**
University Technology Malaysia (UTM), Malaysia
- **Asoke Nath**
St. Xaviers College, India
- **B R SARATH KUMAR**
Lenora College of Engineering, India
- **Binod Kumar**
Lakshmi Narayan College of Technology, India
- **Bremananth Ramachandran**
School of EEE, Nanyang Technological University
- **Dr.C.Suresh Gnana Dhas**
Park College of Engineering and Technology, India
- **Mr. Chakresh kumar**
Manav Rachna International University, India
- **Chandra Mouli P.V.S.S.R**
VIT University, India
- **Chandrashekhar Meshram**
Shri Shankaracharya Engineering College, India
- **Constantin POPESCU**
Department of Mathematics and Computer Science, University of Oradea
- **Prof. D. S. R. Murthy**
SNIST, India.
- **Deepak Garg**
Thapar University.
- **Prof. Dhananjay R.Kalbande**
Sardar Patel Institute of Technology, India
- **Dhirendra Mishra**
SVKM's NMIMS University, India
- **Divya Prakash Shrivastava**
EL JABAL AL GARBI UNIVERSITY, ZAWIA
- **Dragana Becejski-Vujaklija**
University of Belgrade, Faculty of organizational sciences
- **Fokrul Alom Mazarbhuiya**
King Khalid University
- **G. Sreedhar**
Rashtriya Sanskrit University
- **Ghalem Belalem**
University of Oran (Es Senia)
- **Hanumanthappa.J**
University of Mangalore, India
- **Dr. Himanshu Aggarwal**
Punjabi University, India
- **Huda K. AL-Jobori**
Ahlia University
- **Dr. Jamaiah Haji Yahaya**
Northern University of Malaysia (UUM), Malaysia
- **Jasvir Singh**
Communication Signal Processing Research Lab
- **Jatinderkumar R. Saini**
S.P.College of Engineering, Gujarat
- **Prof. Joe-Sam Chou**
Nanhua University, Taiwan
- **Dr. Juan José Martínez Castillo**
Yacambu University, Venezuela
- **Dr. Jui-Pin Yang**

- Shih Chien University, Taiwan
- **Dr. K.PRASADH**
Mets School of Engineering, India
 - **Ka Lok Man**
Xi'an Jiaotong-Liverpool University (XJTLU)
 - **Dr. Kamal Shah**
St. Francis Institute of Technology, India
 - **Kodge B. G.**
S. V. College, India
 - **Kohei Arai**
Saga University
 - **Kunal Patel**
Ingenuity Systems, USA
 - **Lai Khin Wee**
Technischen Universität Ilmenau, Germany
 - **Latha Parthiban**
SSN College of Engineering, Kalavakkam
 - **Mr. Lijian Sun**
Chinese Academy of Surveying and Mapping, China
 - **Long Chen**
Qualcomm Incorporated
 - **M.V.Raghavendra**
Swathi Institute of Technology & Sciences, India.
 - **Madjid Khalilian**
Islamic Azad University
 - **Mahesh Chandra**
B.I.T, India
 - **Mahmoud M. A. Abd Ellatif**
Mansoura University
 - **Manpreet Singh Manna**
SLIET University, Govt. of India
 - **Marcellin Julius NKENLIFACK**
University of Dschang
 - **Md. Masud Rana**
Khunla University of Engineering & Technology,
Bangladesh
 - **Md. Zia Ur Rahman**
Narasaraopeta Engg. College, Narasaraopeta
 - **Messaouda AZZOUZI**
Ziane AChour University of Djelfa
 - **Dr. Michael Watts**
University of Adelaide, Australia
 - **Miroslav Baca**
University of Zagreb, Faculty of organization and
informatics / Center for biomet
- **Mohamed Ali Mahjoub**
Preparatory Institute of Engineer of Monastir
 - **Mohammad Talib**
University of Botswana, Gaborone
 - **Mohammed Ali Hussain**
Sri Sai Madhavi Institute of Science & Technology
 - **Mohd Helmy Abd Wahab**
Universiti Tun Hussein Onn Malaysia
 - **Mohd Nazri Ismail**
University of Kuala Lumpur (UniKL)
 - **Mueen Uddin**
Universiti Teknologi Malaysia UTM
 - **Dr. Murugesan N**
Government Arts College (Autonomous), India
 - **Nitin S. Choubey**
Mukesh Patel School of Technology Management &
Eng
 - **Dr. Nitin Surajkishor**
NMIMS, India
 - **Paresh V Virparia**
Sardar Patel University
 - **Dr. Poonam Garg**
Institute of Management Technology, Ghaziabad
 - **Raj Gaurang Tiwari**
AZAD Institute of Engineering and Technology
 - **Rajesh Kumar**
National University of Singapore
 - **Rajesh K Shukla**
Sagar Institute of Research & Technology-
Excellence, India
 - **Dr. Rajiv Dharaskar**
GH Raison College of Engineering, India
 - **Prof. Rakesh. L**
Vijetha Institute of Technology, India
 - **Prof. Rashid Sheikh**
Acropolis Institute of Technology and Research,
India
 - **Ravi Prakash**
University of Mumbai
 - **Rongrong Ji**
Columbia University
 - **Dr. Ruchika Malhotra**
Delhi Technological University, India
 - **Dr.Sagarmay Deb**
University Lecturer, Central Queensland University,
Australia

- **Saleh Ali K. AlOmari**
Universiti Sains Malaysia
- **Dr. Sana'a Wafa Al-Sayegh**
University College of Applied Sciences UCAS-
Palestine
- **Santosh Kumar**
Graphic Era University, India
- **Sasan Adibi**
Research In Motion (RIM)
- **Saurabh Pal**
VBS Purvanchal University, Jaunpur
- **Seyed Hamidreza Mohades Kasaei**
University of Isfahan
- **Shahanawaj Ahamad**
The University of Al-Kharj
- **Shaidah Jusoh**
University of West Florida
- **Sikha Bagui**
Zarqa University
- **Dr. Smita Rajpal**
ITM University
- **Suhas J Manangi**
Microsoft
- **SUKUMAR SENTHILKUMAR**
Universiti Sains Malaysia
- **Sunil Taneja**
Smt. Aruna Asaf Ali Government Post Graduate
College, India
- **Dr. Suresh Sankaranarayanan**
University of West Indies, Kingston, Jamaica
- **T C.Manjunath**
Visvesvaraya Tech. University
- **T V Narayana Rao**
Hyderabad Institute of Technology and
Management, India
- **T. V. Prasad**
Lingaya's University
- **Taiwo Ayodele**
Lingaya's University
- **Totok R. Biyanto**
Infonetmedia/University of Portsmouth
- **Varun Kumar**
Institute of Technology and Management, India
- **Vellanki Uma Kanta Sastry**
Sreeneedhi
- **Dr. V. U. K. Sastry**
SreeNidhi Institute of Science and Technology
(SNIST), Hyderabad, India.
- **Vinayak Bairagi**
Sinhgad Academy of engineering, India
- **Vitus S.W. Lam**
The University of Hong Kong
- **Vuda Sreenivasarao**
St.Mary's college of Engineering & Technology,
Hyderabad, India
- **Y Srinivas**
GITAM University
- **Mr.Zhao Zhang**
City University of Hong Kong, Kowloon, Hong Kong
- **Zhixin Chen**
ILX Lightwave Corporation
- **Zuqing Zhu**
University of Science and Technology of China

CONTENTS

Paper 1: Estimation of the Visual Quality of Video Streaming Under Desynchronization Conditions

Authors: A.A. Atayero, O.I. Sheluhin, Y.A. Ivanov, A.A. Alatishe

PAGE 1 – 11

Paper 2: A Novel Intra-Domain Continues Handover Solution for Inter-Domain Pmipv6 Based Vehicular Network

Authors: Haidar N. Hussain, Kamalrulnizam Abu Bakar, Shaharuddin Salleh

PAGE 12 – 18

Paper 3: Autonomous Control of Eye Based Electric Wheel Chair with Obstacle Avoidance and Shortest Path Findings Based on Dijkstra Algorithm

Authors: Kohei Arai, Ronny Mardiyanto

PAGE 19 – 25

Paper 4: Eye-based Human Computer Interaction Allowing Phoning, Reading E-Book/E-Comic/E-Learning, Internet Browsing, and TV Information Extraction

Authors: Kohei Arai, Ronny Mardiyanto

PAGE 26 – 32

Paper 5: Very Low Power Viterbi Decoder Employing Minimum Transition and Exchangeless Algorithms for Multimedia Mobile Communication

Authors: Prof. S. L. Haridas, Dr. N. K. Choudhari

PAGE 33 – 36

Paper 6: Outlier-Tolerant Kalman Filter of State Vectors in Linear Stochastic System

Authors: HU Shaolin, Huajiang Ouyang, Karl Meinke, SUN Guoji

PAGE 37 – 41

Paper 7: Handsets Malware Threats and Facing Techniques

Authors: Marwa M. A. Elfattah, Aliaa A.A Youssif, Ebada Sarhan Ahmed

PAGE 42 – 48

Paper 8: Identifying Nursing Computer Training Requirements using Web-based Assessment

Authors: Naser Ghazi, Gitesh Raikundalia, Janette Gogler, Leslie Bell

PAGE 49 – 61

Paper 9: A Comparative study of Arabic handwritten characters invariant feature

Authors: Hamdi Hassen, Maher Khemakhem

PAGE 62 – 68

Paper 10: Pattern Discovery Using Association Rules

Authors: Ms Kiruthika M, Mr Rahul Jadhav, Ms Dipa Dixit, Ms Rashmi J, Ms Anjali Nehete, Ms Trupti Khodkar

PAGE 69 – 74

Paper 11: The macroeconomic effect of the information and communication technology in Hungary

Authors: Peter Sasvari

PAGE 75 – 81

Paper 12: Preprocessor Agent Approach to Knowledge Discovery Using Zero-R Algorithm

*Authors: Inamdar S. A, Narangale S.M., G. N. Shinde**

PAGE 82 – 84

Paper 13: Text Independent Speaker Identification using Integrating Independent Component Analysis with Generalized Gaussian Mixture Model

Authors: N M Ramaligeswararao, Dr.V Sailaja, Dr.K. Srinivasa Rao

PAGE 85 – 91

Paper 14: Energy Efficient Zone Division Multihop Hierarchical Clustering Algorithm for Load Balancing in Wireless Sensor Network

Authors: Ashim Kumar Ghosh, Anupam Kumar Bairagi, Dr. M. Abul Kashem, Md. Rezwan-ul-Islam, A J M Asraf Uddin

PAGE 92 – 97

Paper 15: Eyes Based Electric Wheel Chair Control System- I (eye) can control Electric Wheel Chair -

Authors: Kohei Arai, Ronny Mardiyanto

PAGE 98 – 105

Paper 16: Fuzzy Petri Nets for Human Behavior Verification and Validation

Authors: M. Kouzehgar, M. A. Badamchizadeh, S. Khanmohammadi

PAGE 106 – 114

Paper 17: SVD-EBP Algorithm for Iris Pattern Recognition

Authors: Mr. Babasaheb G. Patil, Dr. Mrs. Shaila Subbaraman

PAGE 115 – 119

Paper 18: Using Semantic Web to support Advanced Web-Based Environment

Authors: Khaled M. Fouad, Mostafa A. Nofal, Hany M. Harb, Nagdy M. Nagdy

PAGE 120 – 129

Paper 19: A Virtual Environment Using Virtual Reality and Artificial Neural Network

Authors: Abdul Rahaman Wahab Sait, Mohammad Nazim Raza

PAGE 130 – 133

Paper 20: Agent based Bandwidth Reservation Routing Technique in Mobile Ad Hoc Networks

Authors: Vishnu Kumar Sharma, Dr. Sarita Singh Bhadauria

PAGE 134 – 139

Paper 21: Sensor Node Deployment Strategy for Maintaining Wireless Sensor Network Communication Connectivity

Authors: Shigeaki TANABE, Kei SAWAI, Tsuyoshi SUZUKI

PAGE 140 – 146

Paper 22: Detection and Extraction of Videos using Decision Trees

Authors: Sk.Abdul Nabi, Shaik Rasool, Dr.P. Premchand

PAGE 147 – 151

Paper 23: An Approach to Improve the Representation of the User Model in the Web-Based Systems

Authors: Yasser A. Nada, Khaled M. Fouad

PAGE 152 – 160

Paper 24: Solving the MDBCS Problem Using the Metaheuristic–Genetic Algorithm

Authors: Milena Bogdanovic

PAGE 161 – 167

Paper 25: Optimized Min-Sum Decoding Algorithm for Low Density Parity Check Codes

Authors: Mohammad Rakibul Islam, Dewan Siam Shafiullah, Muhammad Mostafa Amir Faisal, Imran Rahman

PAGE 168 – 174

Paper 26: A New Approach of Digital Forensic Model for Digital Forensic Investigation

Authors: Inikpi O. Ademu, Dr Chris O. Imafidon, Dr David S. Preston

PAGE 175 – 178

Paper 27: A Data Mining Approach for the Prediction of Hepatitis C Virus protease Cleavage Sites

Authors: Ahmed mohamed samir ali gamal eldin

PAGE 179 – 182

Paper 28: Enhancing Business Intelligence in a Smarter Computing Environment through Cost Analysis

Authors: Saurabh Kacker, Vandana Choudhary, Tanupriya Choudhury, Vasudha Vashisht

PAGE 183 – 190

Paper 29: A Flexible Tool for Web Service Selection in Service Oriented Architecture

Authors: Walaa Nagy, Hoda M. O. Mokhtar, Ali El-Bastawissy

PAGE 191 – 201

Estimation of the Visual Quality of Video Streaming Under Desynchronization Conditions

¹A.A. Atayero, ²O.I. Sheluhin,

^{1,4}Department of Electrical and Information Engineering
Covenant University
Ota, Nigeria

³Y.A. Ivanov, ⁴A.A. Alatishe

^{2,3}Department of Information Security
Moscow Tech. Univ. of Comm. and Informatics
Moscow, Russia

Abstract—This paper presents a method for assessing desynchronized video with the aid of a software package specially developed for this purpose. A unique methodology of substituting values for lost frames was developed. It is shown that in the event of non-similarity of the sent and received sequences because of the loss of some frames in transit, the estimation of the quality indicator via traditional (existing) software are done inaccurately. We present in this paper a novel method of estimating the quality of desynchronized video streams. The developed software application is able to carry out the estimation of the quality of video sequences even when parts of the frame is missing, by means of searching out contextually similar frames and “gluing” them in lieu of the lost frames. Comparing obtained results with those from existing software validates their accuracy. The difference in results and methods of estimating video sequences of different subject groups is also discussed. The paper concludes with adequate recommendations on the best methodology to adopt for specific estimation scenarios.

Keywords—video streaming; encoder; decoder; video streaming quality; PSNR.

I. INTRODUCTION

The assessment of video quality is currently being researched both theoretically and practically. New and objective metrics for estimating the quality of video signals are constantly being developed. These metrics are often in the form of mathematical models that imitate subjective estimates. Calculating Peak Signal-to-Noise Ratio (PSNR) between the source signal and that obtained at the output of the system being analyzed remains the conventional means of estimating the quality of digitally processed video signals using software. There exists a plethora of software solutions from both academia and industry, that allow for estimation of the quality of streaming video; popular examples are: Elecard Video Estimator [1], Video Quality Studio 0.32 [2], MSU Video Quality Measurement Tool [3], PSNR.exe [4].

All the packages cited above are however meant exclusively for the assessment of fluctuations in video information arising because of coding and compression. They cannot process files with frames either lost in the process of transmission or desynchronized vis-à-vis the input sequence.

Video transmission over wireless networks presupposes a possible loss of synchronization between the original video sequence and the decoded copy at the receive end. This is due

to the unpredictable nature of the effect of the transmission medium on the data packet, leading to packet distortion and often times, outright loss of packets. It is for this reason that a manual synchronization between the analyzed video sequences must be done.

The development of a software package that incorporates knowledge of these peculiarities as well as the capability of calculating quality parameters in cases of loss of frames in the video sequence becomes of paramount importance.

II. DESIGNING THE SOFTWARE

A. Theoretical Background

The major peculiarity of pictures is their mode (modal characteristic). There are three possible picture modes namely: Red-Green-Blue (RGB), half-tone scale (formation of pictures using the brightness level) and indexing.

RGB: In this mode, each picture element (pixel) is described by the Red, Green, and Blue color levels. Since any perceivable color can be presented as a combination of these three basic colors, the RGB picture is a full-color picture. Each color is described as an 8-bit information, this allows for the usage of 256 color intensity levels, resulting in 16.7 million colors (i.e. $2^8R \times 2^8G \times 2^8B$), also known as True Color.

Half-tone scale: On the other hand for Half-tone images, each pixel can be described in 8-bit brightness levels ranging from 0 (absolute black) to 255 (maximum brightness). The actual difference between half tone and RGB images lies in the number of color channels: one for half-tone images and three for RGB images. An RGB image may be presented as the superposition of three half-tone images, each of which corresponds to R, G, and B respectively. Three matrices, each of which correspond to one of the RGB colors and determine the pixel color, describe the image. For example, an image of 176x144 pixels would require three matrices of equal size whose elements are the intensity values of the color of each pixel.

The PSNR parameter is a mathematical instrument by which the correspondence (relationship) between the original and distorted video sequences can be established. The greater the difference between the sent and received video sequences, the lower the PSNR value measured in dB, in accordance with the visual logarithmic sensitivity of the human eye.

In the comparison of two video sequences made up of N frames with resolution $D_x \times D_y$ (e.g. for QCIF $x = 144$, $y = 176$; for CIF $x = 288$, $y = 372$) and pixel coordinates $I(n, x, y)$, where $n = 0, \dots, N - 1$, $x = 1, \dots, D_x$; $y = 1, \dots, D_y$ represents the brightness (component Y) of a pixel with coordinates (x, y) in the video frame n), the following equation holds:

$$PSNR_n = 10 \log_{10} \frac{255}{\frac{1}{D_x D_y} \sum_{x=1}^{D_x} \sum_{y=1}^{D_y} [I(n,x,y) - \bar{I}(n,x,y)]^2} \quad (1)$$

Applications used for real-time transmission usually code multimedia information in a standard that is not stringent on packet loss, an example is the popular MPEG coding standard. This standard employs both intra-frame and inter-frame compression with different types of frames (I, P and B). Repeated parts of I, P and B frames are called Group of Pictures (GoP). The choice of the GoP structure affects the properties of the MPEG video; such as file size, which in turn affects the video stream bitrate and ultimately the resultant visual quality. The number and relationship of different types of frames in GoP is chosen relative to coding efficiency for the particular video subject group in question: static, pseudo static, and highly-dynamic (i.e. SSG, PSSG, and HDSG).

The GoP length is a function of the structure and number of frame types used. Short GoP of less than six frames is used when image transition is very fast. If a fast transition of the preceding frame is observed, the presence of a large quantity of P and B frames may worsen the quality. The possible value of P frames is in the range of 2 to 14. Usually, only a small value of P frames is used in practice (say 3 or 4). Only I and P frames (e.g. 1 I-frame and 14 P-frames) can be used if coding occurs with high bitrate (e.g. > 6000 kbps). B-frames guarantee good compression, but like P-frames, they cause degradation of quality in dynamic subjects. A small value of B-frames is used in practice (say 0 or 3).

The most commonly employed GoP structure is IBBPBBPBBPBB. The maximum GoP length according to DVD specification is 18 for NTSC and 15 for PAL. In order to increase coding efficiency, it is important to determine the size and spread of GoP frames. Since the subject of any video sequence may change in time, the GoP structure may also be non-static and change with time. Hence the need for a dynamic determination of the GoP structure. It is possible to automatically determine the GoP structure in real-time with availability of information such as Sum of Absolute Difference (SAD) and Mean of Absolute Difference (MAD). Works on this problem abound in the literature: a relatively easy methodology for determining the GoP structure is developed in [5], [6]. The optimization of GoP structure based on the time relation of sequential frames was considered in [7]. A simple method for determining the GoP structure is given in [8], while [9] presents an algorithm for GoP determination.

The MPEG standard assumes that in most videos, adjacent frames are similar. Since adjacent frames are described based on how they differ from the input frame, assuming that frames within the GoP may be replaced with each other in cases of loss of one of them with negligible effect on the quality of the whole video sequence is logically sound. However, during frame padding, image analysis is necessary to determine the most similar frame for use. The decision to use Matlab for this

purpose was adopted, since it allows for the processing of a large amount of data and has its own scripting language.

III. DESCRIPTION OF DEVELOPED SOFTWARE

The basic principle of the developed program is to estimate distortions added to the video sequence, taking into account frame-wise comparison of the original and received video sequence. PSNR is calculated for each frame of video sequence received (i.e. that passed through the data transmission network). The input data for the program are frames of video sequences (original and received) in bmp format. Each frame is represented as a three-dimensional matrix in the software, while each matrix element can be any value from 0 to 255 that determines the saturation of one of the colors (Red, Green, Blue) of each image pixel. Repeated identical frames in the transmitted and received video sequences are deleted. In cases of different frame numbers in the video sequences, the software will process the minimum value (usually a received sequence). For correct comparison of video sequences, the software compares the received sequences with the original, i.e. synchronizes the video data. Each received frame sequence is compared with the frames in the original sequence within a certain range of frame numbers.



Figure 1. A Method for Frame Synchronization

This interval is set manually with consideration for the adopted video subject group and a certain GoP structure. If the numbers of received and initial frames do not match, this interval is extended to the difference of frames. The principle of frame synchronization is shown in Figure 1. If the search interval is set to be 3 frames, then this value will be increased to 5 frames, because two frames are added due to difference in amounts of the original and received frames in the video sequence (the difference of two frames). Figure 1 shows three cases of frame search, depending on the position of the lost frame in the sequence. An example of when received frames are less than the original is given.

The value of video frame similarity is the average of all matrix elements calculated sequentially for each dimension. To calculate the PSNR, a difference matrix containing the pixel differences of the received and input frames is generated. The data of the difference matrix is also averaged sequentially in each dimension, and the maximum difference in the frames is presented as an 8-bit number ranging from 0 (identical images) to 255 (maximally different frames in terms of comparison of

black to white image). The obtained PSNR values are stored in a file containing the source frame, corresponding number of the received frame and value of PSNR of the frames. The algorithms for the developed software is presented in flow-chart form in Figure 2.

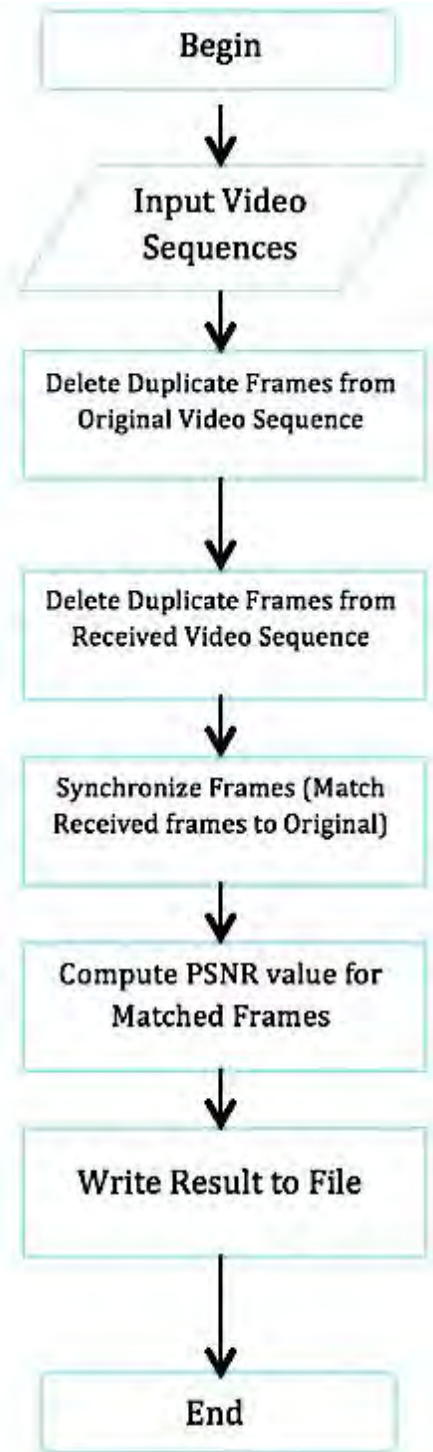


Figure 2. A Method for Frame Synchronization

IV. RESULTS AND DISCUSSION

For the experiment, the video sequences of “Hall”, “Foreman”, and “Football” in YUV format accessible at [10] and recommended for the carrying out test experiment [11] were used. These sequences are characterized by having various subject groups, e.g. static subject group (SSG) - hall, sedentary or pseudo static (PSSG) - foreman, and highly dynamic (HDSG) - football. In evaluating the developed software, two experiments were conducted: In the first experiment simulation of video transmission over a wireless AWGN channel with BER = 10⁻⁴ was done using VCDemo [12]. Synchronization of video sequences was intact. In the next experiment, the de-synchronization of video sequences was achieved by removing frames from the received video using VirtualDub [13] as shown in Figure 3. Frame removal was performed in three different combinations: 1) – the 50th frame was removed; 2) – the 50th to 55th frames were removed; 3) – the 50th to 60th frames were expunged. These removals correspond to 1%, 5% and 10% frames removed.

Quality assessment was performed using the psnr.exe software, a component of the hardware and software complex in detail in [14]. Results of quality assessment of synchronized and desynchronized video sequences are shown in Figure 4a and 4b. Figure 4b shows that the PSNR value after the 50th frame drops sharply and then remains virtually unchanged. However, this is not true, because removing some frames do not affect the next in the long run. The unreliability of this result shows the inability of the conventional software to compare individually the original and received frames, but rather only executes a serial comparison. Thus, it should be noted that for non-correspondence between transmitted and received frames, due to the loss of some of them, the assessment of quality done by existing software is incorrect.

<i>No of Sent Frames</i>	<i>No of Frames Received</i>	<i>PSNR [dB]</i>
47	47	35.14
48	48	36.10
49	49	33.28
56	50	30.65
57	51	29.99
58	52	33.84
59	53	34.71
60	54	54.83
61	55	47.23
62	56	57.49
63	57	57.33

TABLE I. COMPUTATION RESULTS OBTAINED FROM THE PROPOSED SOFTWARE

The purpose of the second experiment was to test and compare the results of the developed software with traditional software in assessing the quality of video sequences.

For these purposes, the original and the tested video sequences were subjected to a frame-by-frame transformation into a format suitable for Matlab processing (bitmap picture .bmp) using VirtualDub [13]. The numbering of frames for each of the three video sequences namely: a) the original, b) distorted with no loss of frame, and c) distorted with frame loss is done sequentially, that is from 1 to 100. The output of the software is a file report in tabular form containing the number of transmitted frames, number of received frames, and the corresponding PSNR in dB (Table 1).



Figure 3. The distortion of the video sequence after transmission over AWGN wireless channel on the 22nd frame (top) and desynchronized video for 51st frame (bottom)

In the absence of lost frames and retention of synchronization, the serial number of the received frame will match that of the transmitted frame. In the event of loss of some frames during transmission, their serial numbers will be missing in the "No of sent frames" column, which means the absence of the corresponding frame in the received sequence. For such frames, the PSNR is not calculated (in this example from 50th to 55th frames). It is therefore necessary to note missing (i.e. lost) frames during analysis and computation of PSNR. Starting from 50th received frame, there is a correspondence with the source frames albeit by a shift of six frames (i.e. a displacement due to the loss of frames 50–55). Evaluation of the accuracy of the calculated values of PSNR by means of the developed software compared to those of traditional (existing) software is presented in Figure 4a. The small observable change is due to conversion in video sequence format from MPEG video to BMP image format.

Figure 4b shows the values of the PSNR indicator calculated using traditional software with and without synchronization. It is shown that in the measurement of PSNR of video sequences from which frames are removed, the traditional software's calculation is incorrect after the 50th frame. The developed software's calculations are accurate, due to the presence of the synchronization function incorporated in it (Figure 4c, 4d).

In Figure 4c a shift in the PSNR graph is observable and it corresponds to the number of lost frames to the left. "null" values may be inserted in place of lost frames, which as a rule describe the worst quality scenario. This corrective approach makes it possible to match the values before and after the loss of frames. Figure 4d shows the insertion of the value PSNR = 20 dB, which characterizes very poor quality. The PSNR values cannot be predicted even under deterministic experimental conditions. We can only indicate the likelihood of quantities taking a certain value or falling within a given interval. However, with knowledge of the probability distribution of this quantity, one can determine its properties and characteristics [14].

Figure 5 shows the quality indicator of video sequences corresponding to different subject groups namely: static (Hall), pseudo-static (Foreman), and highly dynamic (Football). In Figures 5c and 5d show that the calculation of PSNR value using traditional software is incorrect, since some of the results in the area are less than 25 dB, which corresponds to poor quality. If frames are lost in the sequence, PSNR histogram should not change, and there should be a reduction in the number of observations PSNR, i.e. the histogram should only decrease in the vertical axis (number of observations) in those bands, to which the lost frames belonged. This is the reason for the high distortion of the empirical distribution function (Figure 5e).

Figure 6 displays the results of the calculations gotten from the developed software. In Figure 6b as described above, there is a shift in PSNR left by the amount of lost frames. Histograms of PSNR (Figure 6c and 6d) show that the distortions before and after frame removal are virtually identical, indicating the correctness of PSNR calculation after frame removal using the developed software. The distribution functions also vary slightly depending on the number of lost frames (Figure 6e).

Figure 7 shows computation results of developed software and insertion of "null" values. The graphs of PSNR on Figure 7b are identical except for the data number, which equals the quantity of lost frames. The distribution functions also vary slightly depending on the number of lost frames (Figure 7d). Thus, the method of inserting "nulls" does not correspond to the original data for large values of lost frames. One of the methods of mitigating this short fall is by insertion of average PSNR value of video sequences rather than "nulls".

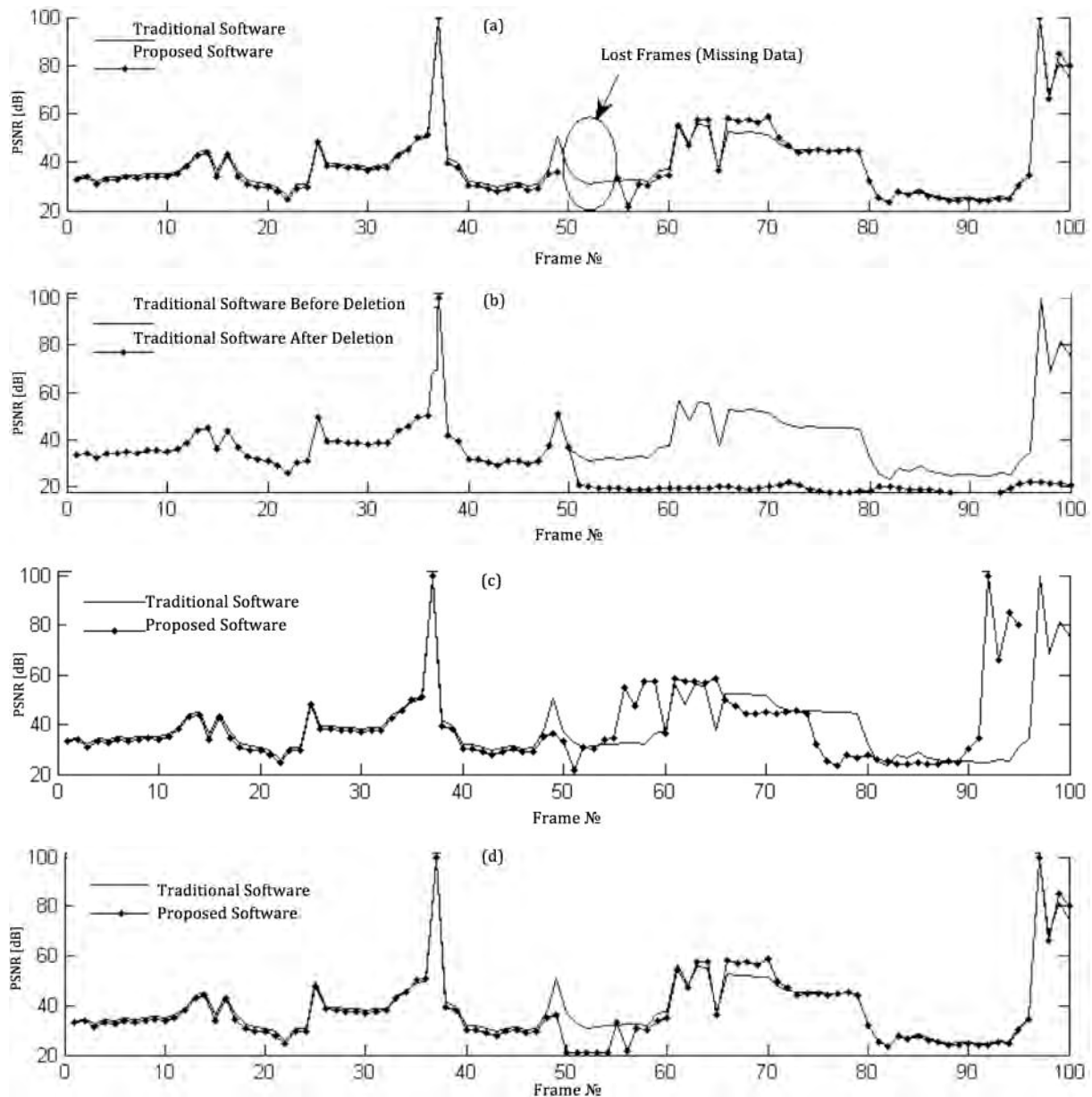


Figure 4. PSNR values for *Forman* video sequence computed with traditional software (without loss of frames) and the proposed software (with loss of frames № 50–55): a) – correspondence of values; b) – Computed values using traditional software; c) – Factual computation of values (observable “gluing” of values in place of lost frames); d) – Insertion of “null” values in place of lost frames.

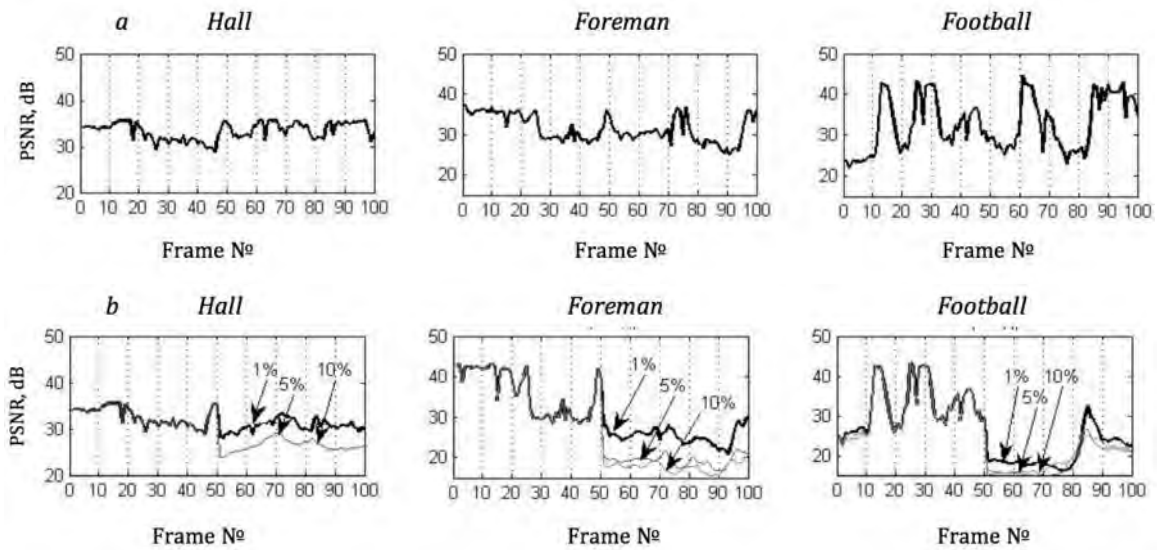
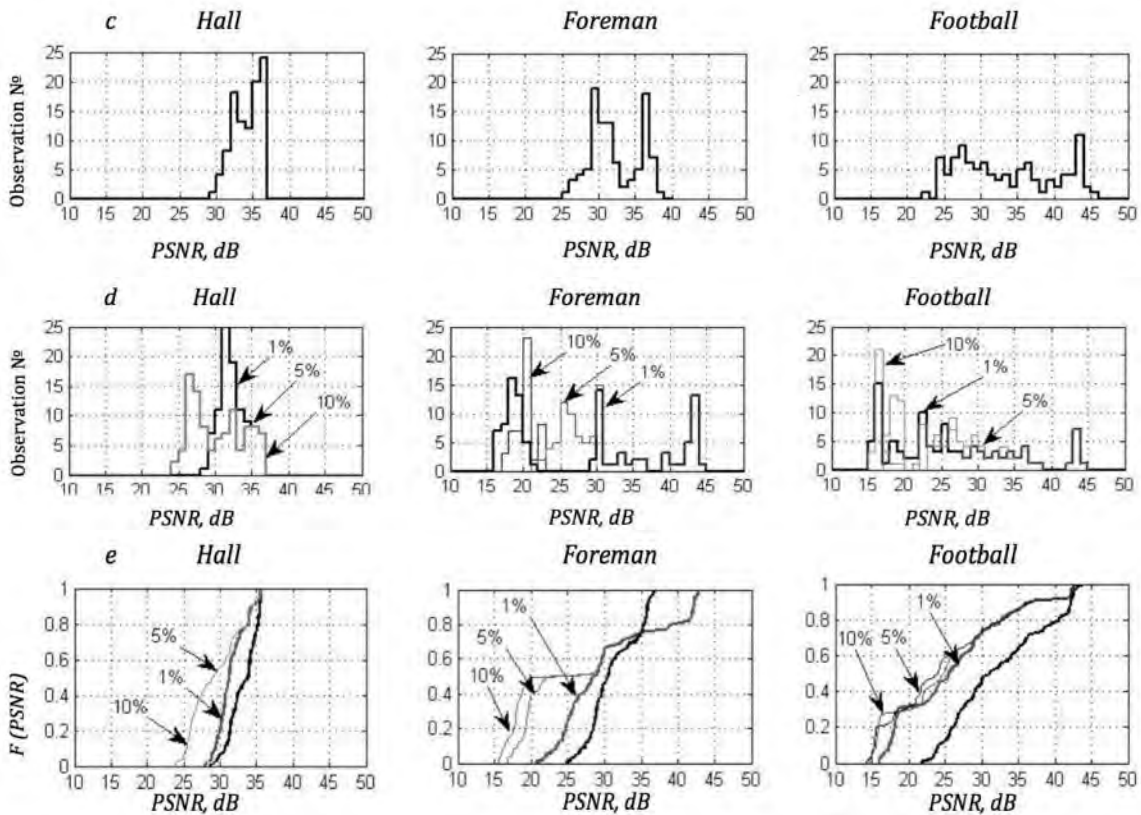


Figure 5a and 5b. PSNR values for video sequences of *Hall*, *Foreman*, and *Football*, distorted during transmission over a wireless channel and computed using traditional software: a) – Synchronized video sequences; b) – for frame loss and desynchronized conditions.



Figures 5c, 5d and 5e. PSNR values for video sequences of *Hall*, *Foreman*, and *Football*, distorted during transmission over a wireless channel and computed using traditional software: c) – Synchronized video sequence distribution histograms; d) – Distribution histograms under frame loss and desynchronization; e) – probability distribution for experimental data with and without frame loss.

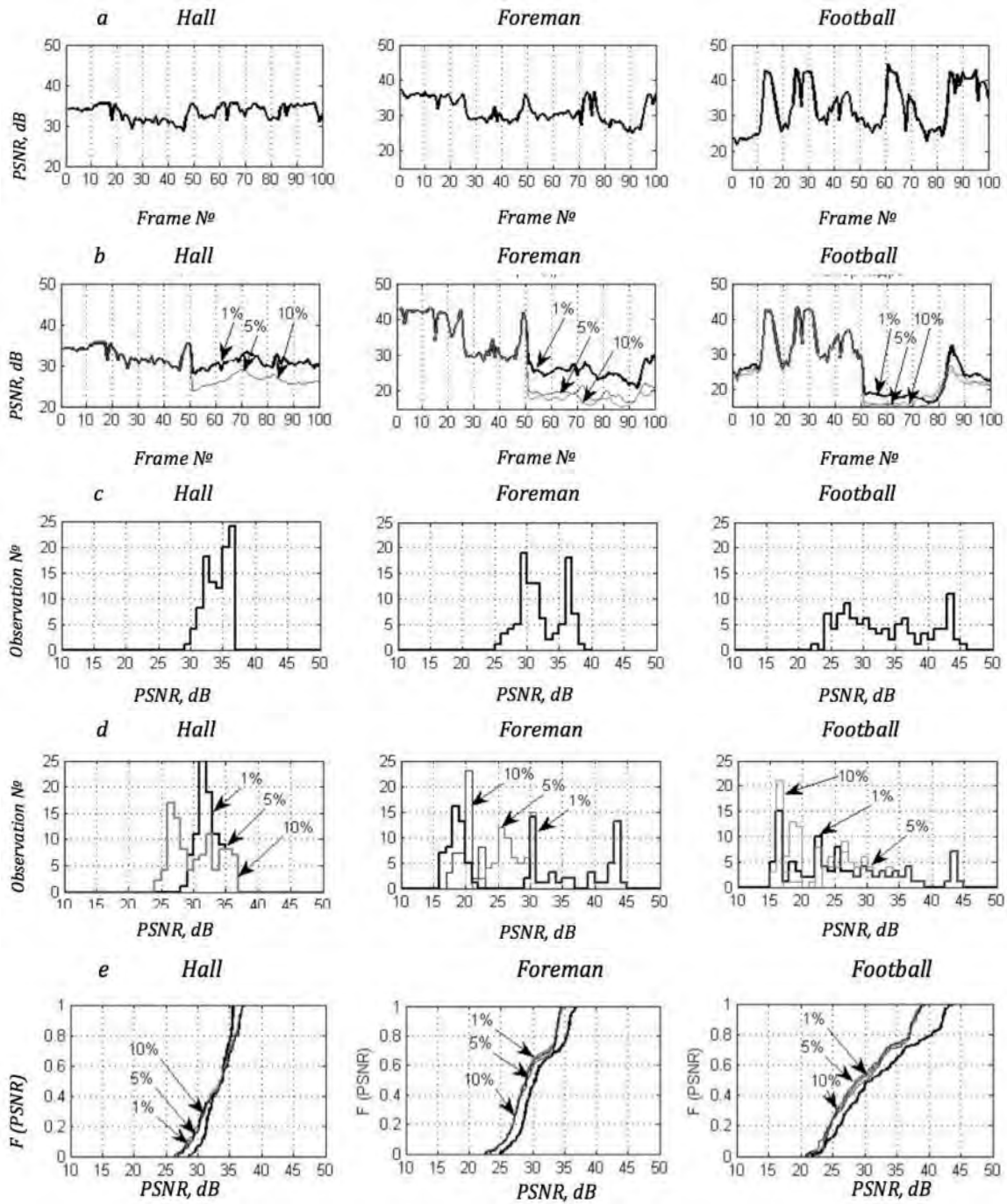


Figure 6. PSNR values for video sequences of *Hall*, *Foreman*, and *Football*, distorted during transmission over a wireless channel and computed using traditional software: a) – Synchronized video sequences; b) – for frame loss and desynchronized conditions; c) – Synchronized video sequence distribution histograms; d) – Distribution histograms under frame loss and Desynchronization conditions; e) Probability distribution of experimental data with and without packet loss.

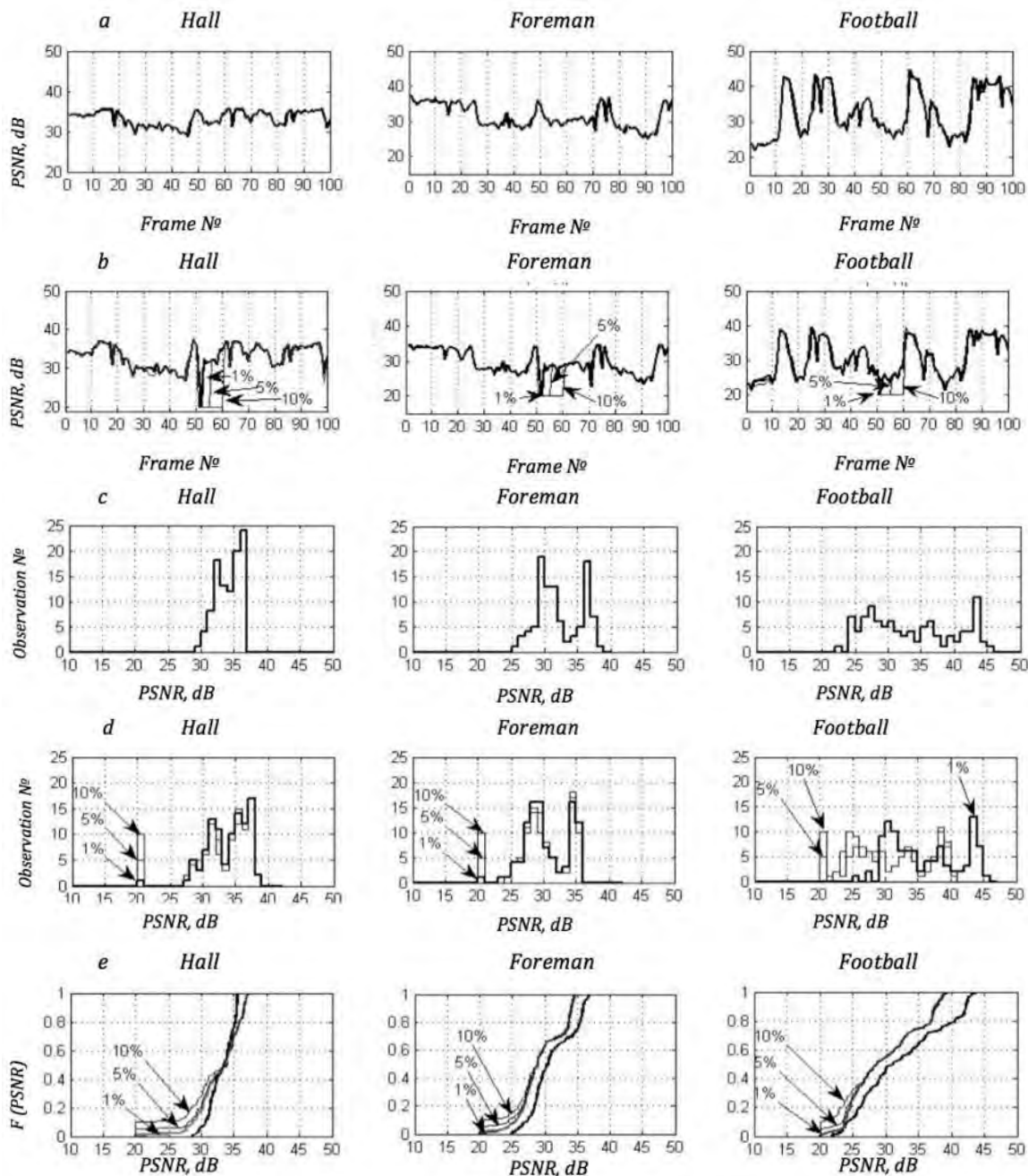


Figure 7. PSNR values for video sequences of *Hall*, *Foreman*, and *Football*, distorted during transmission over a wireless channel and computed using traditional software with insertion of “null” values: a) – Synchronized video sequences; b) – for frame loss and desynchronized conditions; c) – Synchronized video sequence distribution histograms; d) – Distribution histograms under frame loss and desynchronization; e) Probability distribution of experimental data with and without packet loss.

In all, nine experiments were conducted. For ease of presentation of the results, we denote them by their corresponding numbers as follows:

1 – calculated using traditional software with 1% packet loss; 2 – calculated using traditional software with 5% packet loss; 3 – calculated using traditional software with 10% packet loss; 4 – calculated using developed software with 1% packet loss; 5 – calculated using developed software with 5% packet loss; 6 – calculated using developed software with 10% packet loss; 7 – calculated using developed software with 1% packet loss, and insertion of “null” values; 8 – calculated using developed software with 5% packet loss, and insertion of “null” values and 9 – calculated using developed software with 10% packet loss, and insertion of “null” values.

Delta (Δ) values were calculated from the data of distribution functions using formula (2):

$$\Delta = F(\text{PSNR}) - F'(\text{PSNR}) \quad (2)$$

where $F(\text{PSNR})$ – distribution function of PSNR video sequence, calculated with traditional software without lost frames (Figure 4a); $F'(\text{PSNR})$ – distribution function of PSNR video sequence, calculated with the developed software, with a certain number of lost frames (Figure 4c and 4d).

Figure 8 shows the values of $\Delta F(\text{PSNR})$, calculated according to formula (2). Analysis shows that for calculation of desynchronized video sequences using the developed software, a slight difference in the distribution function is noticed as a result of “gluing” values for lost frames. The largest variance from the original distribution function is observed for inserting “null” values. At the same time an increase in the percentage of

lost packets causes a substantial increase in variance with original sequence making this method unsuitable for the calculation of frames, regardless of the video subject group. The histograms of the values of $\Delta F(\text{PSNR})$ are shown in Figure 9.

The mean and variance were gotten from the computed values of $\Delta F(\text{PSNR})$ using the following formula (3). The computation results are presented in Table 2.

$$M = \frac{1}{n} \sum_{i=1}^n \Delta F(\text{PSNR})_i; \quad \sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (\Delta F(\text{PSNR})_i - M)^2 \quad (3)$$

V. CONCLUSIONS AND RECOMMENDATIONS

From the analysis of obtained results, we safely conclude that:

a) Assessing the quality of desynchronized video sequences via traditional software leads to incorrect computational results that do not reflect the true value of PSNR. The error value, $\Delta F(\text{PSNR})$ for traditional software increases from 0.3dB to 0.5dB for SSG and PSSG respectively, and ranges from 0.35dB to 0.45dB for HDSG with increase in percentage of lost frames. The Mean and Variance of the error, $\Delta F(\text{PSNR})$ in estimation of quality using the traditional video software has the highest values, which confirms the incorrectness and inappropriateness of using such software for assessing the quality of results.

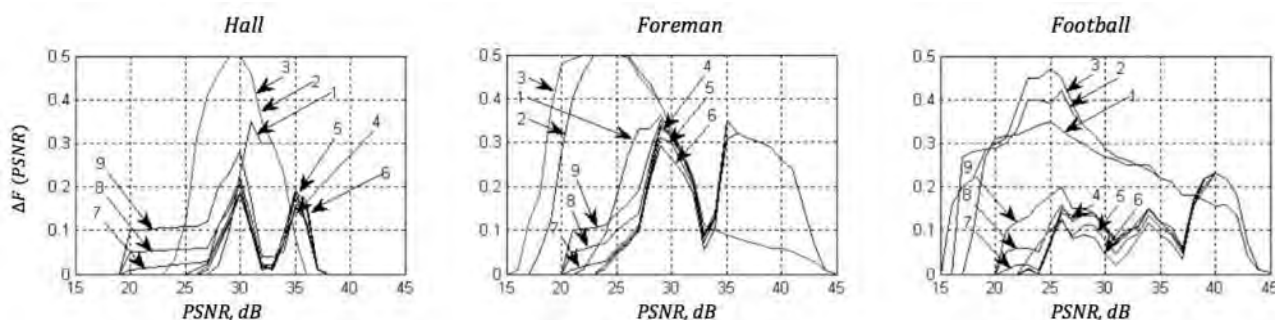


Figure 8. $\Delta F(\text{PSNR})$ values for Hall, Foreman and Football video sequences

TABLE I. MEAN / VARIANCE VALUES FOR $\Delta F(\text{PSNR})$

	Experiment Serial №								
	1	2	3	4	5	6	7	8	9
Hall	0.21	0.33	0.33	0.11	0.11	0.12	0.09	0.12	0.15
	0.014	0.028	0.028	0.004	0.005	0.005	0.003	0.004	0.004
Foreman	0.16	0.25	0.27	0.22	0.22	0.21	0.20	0.21	0.23
	0.014	0.035	0.037	0.013	0.013	0.013	0.014	0.012	0.010
Football	0.25	0.27	0.27	0.14	0.13	0.12	0.13	0.14	0.15
	0.010	0.012	0.015	0.005	0.005	0.004	0.005	0.004	0.003

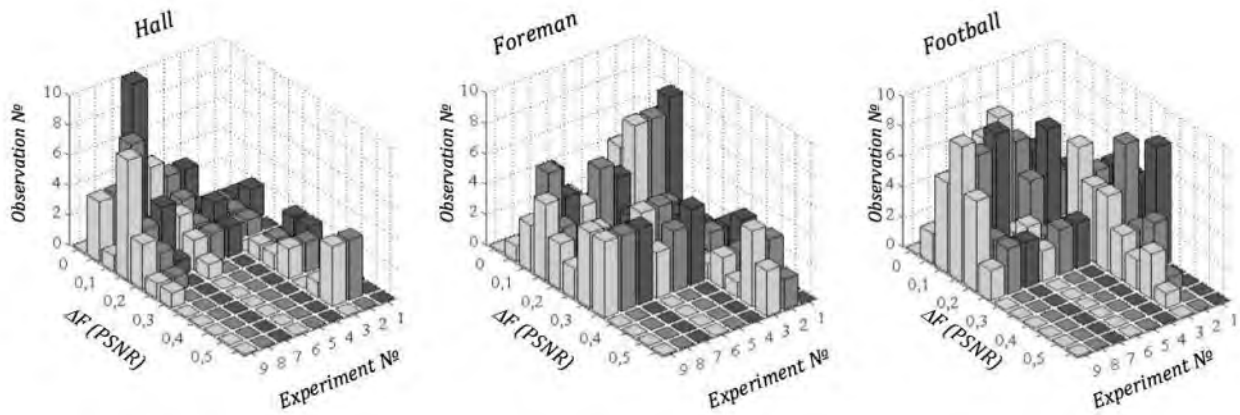


Figure 9. ΔF (PSNR) histograms for Hall, Foreman and Football video sequences

b) Assessing the quality of desynchronized video sequences using the developed software is reliable when used for both normal outcome (“gluing” values instead of lost frames) and for insertion of “null” values. A small difference in the distribution function is observed for inserted values instead of lost frames. Regardless of the percentage of lost frames the error index ΔF (PSNR) has a maximum value of $< 0.2\text{dB}$, $< 0.3\text{dB}$ and $< 0.1\text{dB}$ for the SSG, PSSG and HDSG respectively. In comparison with corresponding values for the traditional software, the superiority of the developed software becomes evident. The biggest difference from the original distribution function is observed in the case of inserted “null” values. With the increase in the percentage of lost packets, the difference in the quality indicator is substantially increased. So, for all subject groups, ΔF (PSNR) increases by an amount not less than 0.05dB and equals 0.25dB , 0.35dB and 0.2dB for the SSG, PSSG and HDSG respectively. Thus, the most appropriate method of estimating the desynchronized video sequences with a small amount of packet loss (less than 10%) in terms of error is by inserting “non-zero” values. The method of inserting “null” values can be recommended for a larger number of lost frames, or if there is a need to restore the number of PSNR values in the original sequence as well. At the same time it is recommended that average values of PSNR be used in lieu of “null” values.

c) It is shown that different subject groups have different effects on the resultant quality of the desynchronized video sequence. For example the PSNR indicator – for SSG is characterized by a slight variation and its histogram has a bimodal shape with a width of 8 dB; – for PSSG has a large variation, with its histogram having a bimodal shape with a width of 14 dB; – for HDSG has the largest variation, its histogram has a bimodal shape with a width of 24 dB. This in turn affects the range of PSNR values containing the error ΔF (PSNR). The highest range of 20dB – 45dB occurs for PSSG and HDSG.

d) Analysis of the mean and variance of the error ΔF (PSNR) shows that the use of insertion is more profitable. Thus with increasing number of lost frames, the mean for the SSG increases from 0.11dB to 0.12dB , and for

PSSG and HDSG decreases from 0.22dB to 0.21dB and from 0.14dB to 0.12dB respectively. Meanwhile, the variance only experiences a slight change. The analysis shows that in the case of a small number of lost frames (1%), the mean of the error ΔF (PSNR) using the method of inserting “null” values has the lowest values (0.09 , 0.2 and 0.13) dB for SSG, PSSG and HDSG respectively. The “glue” method has the largest corresponding to largest (0.11 , 0.22 and 14) dB for each subject group respectively.

e) From the above conclusions, the developed software is recommended for assessing the quality of video streams of different subject groups in desynchronization conditions. With a small number of lost frames (about 1%) it is recommended to use the insertion of “null” values, and the “gluing” method for other cases. This approach guarantees reliable estimation of the quality of the received video stream with minimal estimation error.

VI. SUMMARY

In summary, it is thus shown that in the event of non-similarity of the sent and received frames because of the loss of some frames in transit, the estimation of the quality indicator is done inaccurately using existing software packages. This paper has presented and painstakingly described a novel method of estimating the quality of desynchronized video streams. It is likewise shown that the developed software application is able to carry out the estimation of the quality of a video sequence even when parts of the frame is missing, by means of searching out relevant frames. The accuracy of obtained results is established by comparison with existing software the estimated quality of video streams. The difference in methods of estimating video sequences of different subject groups is also highlighted.

REFERENCES

- [1] Elecard Video Quality Estimator, available at: <http://www.elecard.com/products/products-pc/professional/video-quest/>, Last accessed ...
- [2] Video Quality Studio 0.32, available at: <http://www.visumalchemia.com/vqstudio/#download>, accessed ...
- [3] MSU Video Quality Measurement Tool, available at: <http://www.compression.ru/video/>, accessed ...

- [4] Title http://dvd-hq.info/dvd_compression.php
- [5] Yutaka Yokoyama, "Adaptive GoP Structure Selection for Real-Time MPEG-2 Video Encoding," Proceeding of IEEE International Conference on Image Processing, vol. 2, pp 832-835, Sept. 2000.
- [6] Akio Yoneyama, et.al., "One-pass VBR MPEG encoder using scene adaptive dynamic GoP structure", ICCE2001, pp 174-175, June 2001.
- [7] Xiaodong Gu, and Hongjian Zhang, "Implementing dynamic GOP in video encoding," Proceedings of IEEE International Conference on Multimedia and Expo, vol. 1, pp.349-352, July 2003.
- [8] Akio Yoneyama, et.al., "MPEG Encoding Algorithm with Scene Adaptive Dynamic GOP Structure," IEEE 3rd Workshop on Multimedia Signal Processing, pp 297-302, Sept. 1999.
- [9] Jungwoo Lee, et.al, "Rate-Distortion Optimization Frame Type Selection for MPEG Encoding," IEEE Trans. on Circuits and System for Video Technology, vol. 7, No 3, pp.501-510, June 1997.
- [10] Hall, Foreman, Football video traces, Video Trace Library, Arizona State University, Available at: <http://trace.eas.asu.edu>, Last accessed 2011.11.30.
- [11] ITU-R Recommendation BT.802.-1 Test pictures and sequences for subjective assessments of digital codecs converging signal produced according to Recommendation ITU-R BT.601.
- [12] VCDemo, Signal Information Processing Laboratory, Delft University of Technology, available at: <http://ict.ewi.tudelft.nl/vcdemo>, Last accessed 2011.11.30.
- [13] Virtual dub software, available at: www.virtualdub.org, Last accessed 2011.11.30.
- [14] Atayero A.A., Sheluhin O.I., Ivanov Y.A. and Iruemi J.O., "Effect of Wideband Wireless Access Systems Interference Robustness on the Quality of Video Streaming", Proceedings of the World Congress on Engineering and Computer Science, Vol. II, pp.848-854 WCECS-ICCST 2011, October 19-21, 2011, San Francisco, USA.
- [15] Video Quality Metric software (VQM_pc), Institute for Telecommunication Sciences, Last accessed 2011.12.12, available at: http://www.its.bldrdoc.gov/n3/video/VQM_software_description.php.

AUTHORS PROFILE



Aderemi A. Atayero was born in Lagos, Nigeria in 1969. He graduated from the Moscow Institute of Technology (MIT) with a B.Sc. Degree in Radio Engineering and M.Sc. Degree in Satellite Communication Systems in 1992 and 1994 respectively. He earned a Ph.D in Communications Engineering/Signal Processing from Moscow State Technical University of Civil Aviation, Russia in 2000.

He is a two-time Head, Department of electrical and Information Engineering, Covenant University, Nigeria. He was the coordinator of the School of Engineering of the same University.

Dr. Atayero is a member of a number of professional associations including: the Institute of Electrical and Electronic Engineers, *IEEE*, the International Association of Engineers, *IAENG*, among others. He is a registered member of the Council for the Regulation of Engineering in

Nigeria, *COREN*. He has a number of scientific publications in International peer-reviewed journals, proceedings, and edited books. He is on the editorial board of a number of highly reputed International journals. He is a recipient of the '2009/10 Ford Foundation Teaching Innovation Award'. His current research interests are in Radio and Telecommunication Systems and Devices; Signal Processing and Converged Multi-service Networks.



Oleg I. Sheluhin was born in Moscow, Russia in 1952. He obtained an M.Sc. Degree in Radio Engineering 1974 from the Moscow Institute of Transport Engineers (MITE). He later enrolled at Lomonosov State University (Moscow) and graduated in 1979 with a Second M.Sc. in Mathematics. He received a PhD at MITE in 1979 in Radio Engineering and earned a D.Sc. Degree in *Telecommunication Systems and Devices* from Kharkov Aviation Institute in 1990. The title of his PhD thesis was '*Investigation of interfering factors influence on the structure and activity of noise short-range radar*'.

He is currently Head, Department of Information Security, Moscow Technical University of Communication and Informatics, Russia. He was the Head, Radio Engineering and Radio Systems Department of Moscow State Technical University of Service (MSTUS).

Prof. Sheluhin is a member of the International Academy of Sciences of Higher Educational Institutions. He has published over 15 scientific books and textbooks for universities and has more than 250 scientific papers. He is the Chief Editor of the scientific journal *Electrical and Informational Complexes and Systems* and a member of Editorial Boards of various scientific journals. In 2004 the Russian President awarded him the honorary title '*Honored Scientific Worker of the Russian Federation*'.



Yury A. Ivanov was born in Moscow, Russia in 1985. He obtained an M.Sc. degree in Systems, network and devices in telecommunications from Chuvash State University in 2007. He obtained a Ph.D in Telecommunication Networks and Systems in 2011 from Moscow State University of Communication and Informatics. His dissertation topic was '*The impact of errors in channels of broadband wireless access systems on the quality of streaming H.264/AVC video*'.

Dr. Ivanov has published over 35 scientific papers and his current research interests include Radio and Telecommunications Systems and Devices; transmission of multimedia data across telecommunication networks; assessment of the quality of video sequences.



Adeyemi S. Alatishe was born in Lagos, Nigeria in 1986. He graduated from Covenant University (CU) with a B.Eng. Degree in Information and Communication Technology in 2008 and is currently pursuing a Masters of Engineering (M.Eng) in Information and Communication Technology at the same University as a Research Assistant in the Department of Electrical and Information Engineering.

Mr. Alatishe is a member of a number of professional associations including: the Institute of Electrical and Electronic Engineers, *IEEE*, the International Association of Engineers, *IAENG*, among others. His current research interests are in Satellite communications, Digital signal processing and control systems, E-learning and Operations research.

A Novel Intra-Domain Continues Handover Solution for Inter-Domain Pmipv6 Based Vehicular Network

Haidar N. Hussain, Kamalrulnizam Abu Bakar
Faculty of Computer Science & Information Systems
Universiti Teknologi Malaysia (UTM)
Johor, Malaysia

Shaharuddin Salleh
Faculty of Science Dept. Of Mathematics
Universiti Teknologi Malaysia (UTM)
Johor, Malaysia

Abstract— IP mobility management protocols (e.g. host based mobility protocols) incur significant handover latency, thus aggravate QoS for end user devices. Proxy Mobile IPv6 (PMIPv6) was proposed by the Internet Engineering Task Force (IETF) as a new network-based mobility protocol to reduce the host based handover latency. However the current PMIPv6 cannot support the vehicles high mobility while the vehicles motion within PMIPv6 domain. In this paper we introduce a novel intra-domain PMIPv6 handover technique based vehicular network using Media Independent Handover (MIH). The novel intra-domain PMIPv6 handover based vehicular network improves the handover performance of PMIPv6 by allowing the new PMIPv6 domain to obtain the MIIS information to estimate whether the handover is necessary or not before the vehicles movement to the second MAG of the new PMIPv6 domain. We evaluate the handover latency and data packet loss of the proposed handover process compared to PMIPv6. The conducted analysis results confirm that the novel handover process yields the reduced handover latency compared to that of PMIPv6 and also prevents data packet loss.

Keywords- PMIPv6; MIH; MIIS.

I. INTRODUCTION

The advanced wide deployment of wireless technologies in next generation networks laid the foundation stone of vehicular communication allowing vehicles to connect with the internet during movement. However, the vehicle will experience service interruption during the handover process, i.e., vehicle change its attachment point while maintaining ongoing sessions, due, to distinguishable handover latency and data packet loss.

Recently the developed PMIPv6 network-based protocol [5] allows an unaltered and mobility-unaware vehicle to change its attachment point while maintaining network connectivity. Compared with previous developed host-based mobility management protocols such as Mobile IPv6, Fast Mobile IPv6 (FMIPv6), and Hierarchical Mobile IPv6 (HMIPv6) [9] normally these host-based mobility protocols that are introduced by the Internet Engineering Task Force (IETF) for IP-based networks focus mainly on Mobile IP (MIP) [7]. IETF has developed both MIPv4 [6] and MIPv6. The importance of MIPv6 development is to support one of the most important requirements, which is the efficient support of mobility to provide continuous connectivity.

PMIPv6 is a desirable mobility management protocol designed for telecommunication service providers as well as

manufacturers. As PMIPv6 is deployed in mobile networks, manufacturers do not need to implement a mobility stack for the vehicle. From the telecommunication aspect, they can easily manage control the mobility services [12] Moreover, from the viewpoint performance, PMIPv6 generally outperforms the developed host-based mobility management protocols [8].

Accordingly, the investigation and analysis on the handover process of PMIPv6, which is expected to be a base mobility management protocol for next generation wireless networks (NGWN), is required. In addition, a proposal for improving the handover performance is a desirable work. Nevertheless, in recently published literatures [8] [14], improvements for the handover process of PMIPv6 have not been considered. In this paper, we introduce a novel intra-domain handover solution support for vehicles that cross the inter-domain PMIPv6 network (two different PMIPv6 networks) as far as our concern this is the first intra-domain PMIPv6 based vehicular ad-hoc networks research work done in PMIPv6 area.

The rest of this paper is organized as follows: section II introduces current PMIPv6 handover process. In section III we introduce the novel intra-domain handover solution design for solving the disconnection problem when roaming between two PMIPv6 domains and switching between to MAGs in the new LMA domain. In section IV we develop an analytical model to evaluate our novel intra-domain handover solution. Then, in section V, we present the conducted analysis result. Finally, section VI, concludes this paper.

A. Proxy Mobile ipv6 (PMIPv6)

The Internet Engineering Task Force designed Proxy Mobile IPv6 (PMIPv6) to support network-based IP mobility management for MNs, without requiring its involvement in any related IP-mobility functions. Mobility management in PMIPv6 is provided to MN irrespective of the presence or absence of Mobile IPv6 functionality [6]. Fig 1 shows the conceptual PMIPv6 domain.

PMIPv6 extends the signaling of MIPv6 and reuses most of MIPv6 concepts such as HA functionality. In addition it introduces two new elements known as Local Mobility Anchor (LMA) and Mobile Access Gateway (MAG) [4]. The LMA behaves similar to the HA in MIPv6 in the PMIPv6 domain and also it introduces additional capabilities required for network-based mobility management [2].

PMIPv6 supports the MN within a topological localized domain by utilizing the MAG entity. MAG organizes with the

access routers and handles mobility signaling on behalf of the MN.

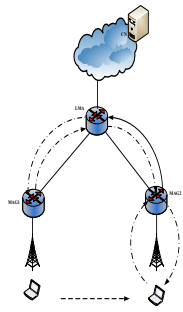


Figure 1. PMIPv6 network architecture

PMIPv6 protocol operation consists of four phases. In the first phase, MAG retrieves the MN's profile using its current identifier. The Binding Update (BU) is the second phase, in which the MAG sends a Proxy Binding Update (PUB) request to the LMA in order to register the current point of attachment of the MN. Accordingly, a binding cache entry and a tunnel for the MN's home network prefix will be created. The third phase will be the MAG emulating the mobile node's home interface on the access interface. Therefore, the MN will always believe it is in the home network. Fourthly, the LMA replies with a Proxy Bind Acknowledge (PBA) message with the MN's HNP. After receiving the Router Advertise (RA) message, the MN configures its IP address by using the contained prefix. For packet routing, the LMA is able to route all received packets over the established tunnel to the MAG. The MAG forwards these packets to the MN. Additionally, the MAG will relay all the received packets over the tunnel to the LMA and then they will be routed towards the CN. Fig 2 shows the procedure when a MN joins a PMIPv6 domain.

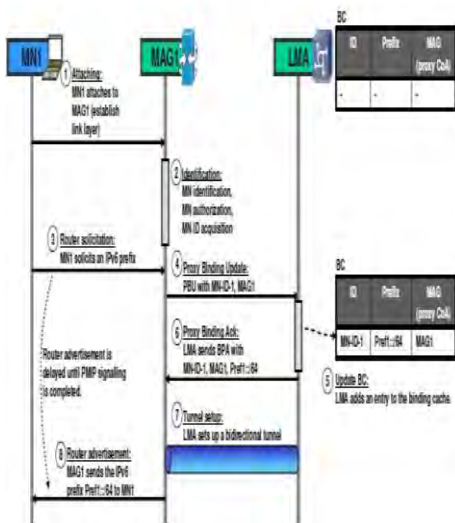


Figure 2. the procedure of a MN joining a PMIPv6 domain.

While the MN is roaming in the PMIPv6 domain, the protocol ensures that the MN is eligible to obtain its home address on any access link [2] on condition that it is roaming in

the same PMIPv6 domain. That is, that the serving PMIPv6 assigns a unique home network prefix, Pre-MN-Prefix, to each MN and this prefix conceptually follows the MN were ever it moves within the PMIPv6 domain [3].

As a result there is no need to perform address configuration to reconfigure a new address for the MN every time it changes its point of attachment. This in turn, optimizes handover performance by reducing the latency that is caused because of the address configuration. Also, because of the MAG network element which performs the network signaling on behalf of the MN, PMIPv6 reduces the binding update delay by reducing the round trip time, thus reducing handover latency. The procedure of the attachment is explained as follows:

- a) Attaching:
MN1 attaches to the MAG1 through a point-to-point link and establishes the link layer. Any access technology is possible provided that it emulates a point-to-point behavior (e.g. PPP, PPPoE).
- b) Identification:
MAG1 authenticates MN1 based on its link layer address (e.g. MAC address) and ascertains what MN1 is permitted to do (authorization). The authorization step may use existing services like LDAP or RADIUS.
- c) Router solicitation:
MN1 sends a router solicitation to obtain an IPv6 prefix. MAG1 will not send a router advertisement until it obtained a prefix for MN1 from the LMA (step 6, PBA).
- d) Proxy binding update (PBU):
MAG1 sends a proxy binding update to the LMA. This PBU associates the MAG1 address with the identity of the MN.
- e) Allocate prefix, update BC:
The LMA allocates a prefix for MN1 (Home Network Prefix). The LMA creates an entry in its BC. The entry contains the MN1 ID (MN-ID-1), the address MAG1 of the proxy MAG (proxy-CoA) as well as the prefix assigned to MN1.
- f) Proxy binding ACK (PBA):
The LMA sends a PBA back to MAG1. The PBA contains the information of the BPC entry created in step 5.
- g) Tunnel setup:
The LMA and MAG1 establish a bidirectional IPv6-in-IPv6 tunnel that is used for tunneling packets to and from MN1. The LMA sets a route through the tunnel for traffic that is addressed to the MN.

B. Media Independent Handover IEEE 802.21

Media Independent Handover (MIH) specification is primitive to provide link layer information and other related information to the upper layers to optimize handover. Fig 3 shows MIH architecture. One of the most important categories is the Event Service (MIES); it is used for the Layer 3 handover hints. Events provide the data link layer conditions to layer 3 or

reflect the response of Layer 3. The representative event primitives include link going down, link down, and link going up. Layer 3 handover occurs when the vehicle changes its point of attachment after an inter-domain movement (inter-network or intra-foreign-network-movement). Inter-domain movement means that the vehicle changes its attachment point between two different networks. When layer 3 handover occurs, the vehicle will lose its IP-connectivity; it hence will lose the ongoing transmission. In addition MIH provides Command Services (MICS); it includes a set of commands that might be sent from higher layers to lower layers, and Information Services (MIIS); provides a set of information, including query/response structure to allow vehicles to discover and obtain information about available networks in a geographical area, attachment capabilities/ network point of attachment, and other information related to the network.

MIH provides these services to MIH user Service Access Point (SAP). MIH-SAP and MIH-Link-SAP serve as MIHF interfaces to L3 and above layers, in addition to lower layers (L2 and below) respectively.

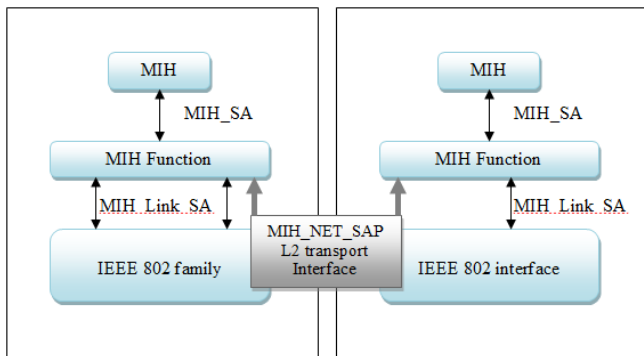


Figure 3. MIH architecture

II. NOVEL INTRA-DOMAIN PMIPv6 HANDOVER TECHNIQUE

We extend the proposed inter-domain PMIPv6 handover scheme in [1] to support intra-domain PMIPv6 by defining new MIH primitives and parameters. As shown in table 1 and Fig 4 below; a new primitive is introduced in our proposed handover scheme known as “MIH-Prefix Info” (Media Independent Handover Prefix-Information).

The MIH-Prefix Info will contain information about the current serving PMIPv6 network domain. The stored information pulled by MIHF of the serving MAG represents lower and upper layer (e.g. L2, L3) information of the serving PMIPv6 domain. MAG pulls the information from the vehicle using MIH-links list and MIH-link Available. In this novel intra-PMIPv6 handover technique we suggest that the “Prefix” parameter is added to these primitives and thereby the serving MAG gets the information by pulling and using the prefix. Fig 5 shows the proposed handover mechanism using MIH.

We propose a Homogeneous Network Information (HNI) Container and it is assumed that the HNI is embedded within the vehicle, Logical interface and is connected with a PMIPv6 domain using MIH (IEEE 802.21) services PMIPv6 architecture that is shown in Fig 4.

TABLE I. NEW PROPOSED PRIMITIVES AND PARAMETERS

Primitives	Service	Parameters
MIH-PrefixInfo	CS	Interface ID, Prefix
MIH-Link List	IS	Interface ID, Prefix, MAC Address, BW, Quality Level
MIH-LinkAvailable	ES	Interface ID, Prefix, MAC Address, BW, Quality Level
MIH-LinkGoingDown	ES	Interface ID, MAC Address, BW, Quality Level
MIH-LinkDown	ES	Interface ID, Prefix, MAC Address
MIH-LinkUp	ES	Interface ID, Prefix, MAC Address

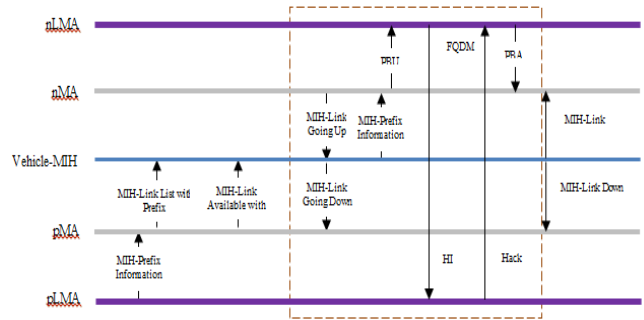


Figure 4. proposed handover mechanism using MIH

HNI defined to facilitate the storage and information retrieval of the L2 and L3 static information of current and neighboring networks obtained through the IEEE 802.21 MIIS. The IE known as the ‘MIH-PrefixInfo’ is used to provide prefixes information about the current network and neighboring networks. Alongside with L2 information, they form the proposed pre-defined (HNI) container. Hence, the information will be used by the LMA once it receives information about the vehicles link status based on this information the LMA will generate a bi-directional tunnel with the next new MAG (i.e. MAG 2 in the new domain) as shown in Fig 6. In addition we propose a logical layer embedded within layer 3 to help support reducing the handover latency by allowing the vehicle to connect with MAG (i.e. MAG2) seamlessly while maintaining it active connection with the first nMAG (i.e. MAG1).

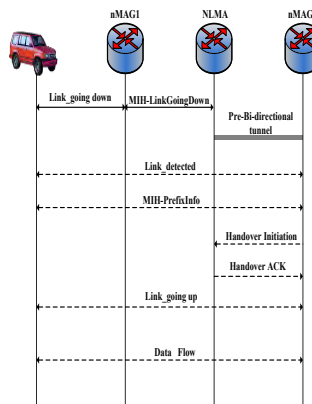


Figure 5. novel intra-domain PMIPv6 handover signaling

We summarize the process of our proposed intra domain scheme as listed below:

- When the serving MAG senses that the vehicle is about to disconnect a link going down event, it will inform the LMA of this event. MAG is enabled to trigger the link event using MIH-Link going down.
- LMA by receiving this indication from the serving MAG will pre-establish a tunnel with the next serving MAG (we assume that the vehicle moves in one direction).
- Once the next MAG (nMAG2) senses the vehicles signal within the overlap area it will pull HNI container and will send it to the LMA. Then nMAG2 will emulate the vehicles to its connection link by establishing the connection through the logical interface.
- The serving LMA will remove the tunnel with the pMAG and the packets will be forwarded through the new path.

By using the logical interface LMA will not buffer the packets until the handover with the next nMAG is completed but it will directly send the packets to the vehicle through the logical interface.

III. ANALYTICAL MODEL AND ANALYSIS

A. Network Model

Fig 6 shows the network model, which includes vehicle, BS, mobile access gateway (MAG), local mobile anchor (LMA), and correspondent node (CN). There are two LMA domains and each LMA has n MAGs. The coverage of LMA is called domain, and the coverage of BS is known as cell. In other words, each domain has n cells. A BS connected to a MAG has a wireless interface for connecting vehicle (s). In this paper, we suppose that the vehicle has moved for inter-domain PMIPv6 to intra-domain PMIPv6 (in this case the process will be an intra-domain process whereby the vehicle is roaming within the same LMA domain).

Furthermore, we assume that once the vehicle passes the overlap area the nLMA will be able to intelligently calculate the stay time of the vehicle within the communication range of the first serving MAG based on the MIH information.

We adopt the vehicle mobility model in where the direction of the vehicle motion in an LMA domain is uniformly distributed on $[0, 2\pi]$.

For simplicity, we assume that the shape of the coverage area of a MAG is circular (non-circular areas, such as hexagonal shaped areas, can be reasonably approximated with the same size) and an intra-PMIPv6 consists of n MAGs with the same size of the coverage area of aS_{AR} . Vehicle (s) move at an average velocity of V .

Let μ_c, μ_d be cell crossing rate and domain crossing rate, respectively. Furthermore, let μ_l be the cell crossing rate for that vehicle which is within the same PMIPv6 domain.

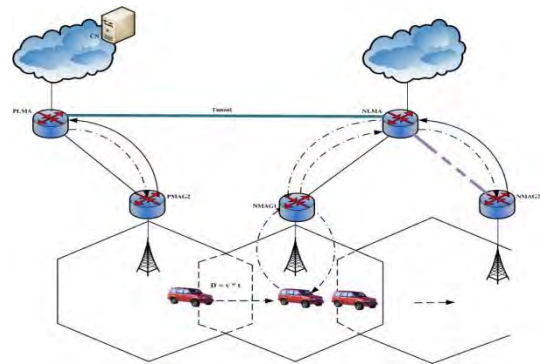


Figure 6. proposed network architecture

Assuming that each AR has a coverage area of S_{AR} , the border crossing is given by [13].

$$\mu_c = \frac{2v}{\sqrt{\pi S_{AR}}} \quad (1)$$

$$\mu_d = \frac{2v}{\sqrt{\pi n S_{AR}}} = \frac{\mu_c}{\sqrt{n}} \quad (2)$$

$$\mu_l = \mu_c - \mu_d = \frac{\sqrt{n-1}}{\sqrt{n}} \mu_c \quad (3)$$

The residence time in a cell and in a domain follows exponential distribution with parameters μ_c and μ_d , while session arrival process follows a Poisson distribution with rate λ_s . Hence, the average number of cell crossing and domain crossing can be obtained as follows:

$$E(N_c) = \frac{\mu_c}{\lambda_s}, \quad (4)$$

$$E(N_d) = \frac{\mu_d}{\lambda_s}. \quad (5)$$

From (2), and (5), we obtain:

$$E(N_d) = \frac{1}{\sqrt{n}} \frac{\mu_c}{\lambda_s} \quad (6)$$

Let $E(N_l)$ be the average number of cell crossing rate of a vehicle, which is in the same PMIPv6 domain (intra-domain). The expression will be as follows

$$E(N_l) = \frac{\mu_l}{\lambda_s} \quad (7)$$

Crossing between to LMA domains, a vehicle crosses between two subnets. Thus, from (4) and (5), the average number of intra-domain handover $E(N_s)$ is given by (8).

$$E(N_s) = E(N_c) - E(N_d) = \frac{\mu_c}{\lambda_s} - \frac{1}{\sqrt{n}} \frac{\mu_c}{\lambda_s} = \frac{\mu_c}{\lambda_s} \left(1 - \frac{1}{\sqrt{n}}\right) \quad (8)$$

For a system that follows a Poisson Process, the delay experienced by packets that are transmitted within the network will be equaled to the transmission delay and the propagation time delay at every link in the network. The network links are combination of both wired and wireless links with the failure probability of P_f .

B. Parameter Analysis

We analyze the important performance metrics such as handover latency $L_H^{(.)}$ which is defined as the time interval during the vehicle cannot receive any data packets while performing the handover process. The notations used for our analysis are as introduced in [9] as listed in table 2:

TABLE II. TABLE 2 THE NOTATIONS USED IN THE ANALYTICAL MODELING

T_{L2}	Link-layer handover latency, which mainly depends on an implementation chipset for a wireless interface
T_{WRS}	The random amount of delay before sending an initial RS message
T_{RS}	The arrival delay of the RS message sent from the MN to the MAG at the new access network
$T_{LU}^{(PMIPv6)}$	The delay of the location update (registration) for the MN
$T_P^{(PMIPv6)}$	The arrival delay of the first packet sent from the LMA to the MN
$T_P^{(PRO)}$	is the arrival delay of the first packet sent from the MAG to the MN
T_{RS}	The buffering start time
T_{RE}	The buffering end time
$T_{L2-REP}^{(PRO)}$	The arrival delay of the L2 report sent from the MN to the MAG
$T_{DREG}^{(PRO)}$	The delay of the de-registration for the MN
$T_{HNPP}^{(PRO)}$	The delay of the HNPP (Home network Prefix process)
$M_S^{(X)}$	The size of the message X, which is used in the handover i.e., $X \in \{PBU, RS, HD, DATA\}$

1) Handover latency of PMIPv6

Despite the handover timing diagram of PMIPv6 handover process shown in Fig 7. Let $L_H^{(PMIPv6)}$ represents the handover latency of PMIPv6 handover process. Then, $L_H^{(PMIPv6)}$ will be expressed as:

$$L_H^{(PMIPv6)} = T_{L2} + T_{WRS} + T_{RS} + T_{LU}^{(PMIPv6)} + T_P^{(PMIPv6)} \quad (9)$$

Where

T_{WRS} is determined as a value between 0 and $MAX_RTR_SOLICITATION_DELAY$ (Narten & T., Nordmark, E., & Simpson W., 1998). It is assumed that T_{WRS} is uniformly distributed in the interval $[0, MAX_RTR_SOLICITATION_DEELAY]$.

T_{RS} is calculated as:

$$T_{RS} = \left(\frac{M_S^{(RS)}}{b_{WL}} + t_{WL} \right) + \left(\frac{M_S^{(RS)}}{b_{WL}} + t_{WL} \right) * \sum_{n_f}^{\infty} n_f * \text{Prob}\{n_f \text{ failures and 1 success}\}, \quad (10)$$

Where $M_S^{(RS)}$ is the RS message size. b_{WL} is the bandwidth of the wireless link between MN and MAG. t_{WL} is the wireless link propagation time. n_f is the number of message failures over the wireless link. Suppose P_f is the link failure probability. By applying P_f in equation (10), it is rewritten as [10]:

$$T_{RS} = \left(\frac{M_S^{(RS)}}{b_{WL}} + t_{WL} \right) + \left(\frac{M_S^{(RS)}}{b_{WL}} + t_{WL} \right) * \frac{P_f}{1-P_f} \quad (11)$$

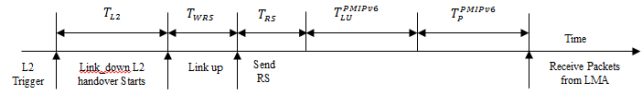


Figure 7. PMIPv6 Handover timing diagram

It is further assumed that the wired link between LMA and MAG is robust and no message failure is expected. Then $T_{LU}^{(PMIPv6)}$ in equation 1 is obtained as:

$$T_{LU}^{(PMIPv6)} = n_h \left(\frac{M_S^{(PBU)}}{b_{WD}} + t_{WD} \right), \quad (13)$$

Where $M_S^{(PBU)}$ is the size of the PBU message. n_h is the number of link hops between the MAG and LMA. b_{WD} is the bandwidth of the wired link between the MAG and LMA. t_{WD} is the propagation time for the wired link. And, $T_P^{(PMIPv6)}$ is obtained by:

$$T_P^{(PMIPv6)} = n_h \left(\frac{M_S^{(HD)} + M_S^{(DATA)}}{b_{WD}} + t_{WD} \right) + 2 \left(\frac{M_S^{(DATA)}}{b_{WL}} + t_{WL} \right) * \frac{P_f}{1-P_f} \quad (14)$$

Where $M_S^{(HD)}$ and $M_S^{(DATA)}$ are the sizes of IPv6 header and messages, respectively. Recall data packets for the MN are traversed via the bi-directional tunnel established between the LMA and new MAG (i.e. MAG2). In addition the LMA knows about the MN attachment with the new MAG as it receives the new MAG PBU message.

2) Novel Intra-domain PMIPv6 solution

Depicts the handover timing diagram of the intra-domain PMIPv6 handover technique the handover process is shown in Fig 8.

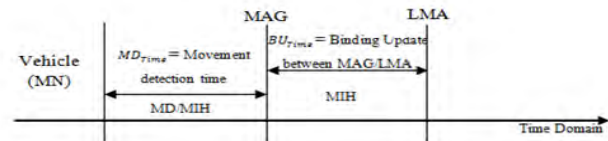


Figure 8. Handover latency time domains of the proposed Inter-domain PMIPv6

Let $L_H^{(P \text{ intra-domain PMIPv6})}$ be the handover latency for the proposed intra-domain PMIPv6 handover latency. Then, it is expressed as:

$$L_H^{(\text{intra-domain PMIPv6 Solution})} = T_{L2} + 2 * T_{\text{vehicle,MAG}} \quad (15)$$

Furthermore, the processing time for the novel intra-domain PMIPv6 is calculated as shown below:

$$T_{L2} + 2 * T_{\text{vehicle,MAG}} + n_h \left(\frac{M_S^{(PBU)}}{b_{WL}} + t_{WL} \right) \quad (16)$$

In equation (16), the data transmission delay over the wired link between LMA and the MAG is excluded, which has taken account in the basic handover process as shown in equation (14). This, is because the proposed intra-domain PMIPv6 handover process allows the nLMA to establishes a bi-directional tunnel with the next nMAG1 that falls in the vehicle movement direction; while the vehicle is within the overlap area of the first new nMAG1 and second nMAG2 within the nLMA. Then the LMA forwards the packets destined to the vehicle through the next nMAG2. Accordingly, as soon as the vehicle attaches to the MAG, it will receive the data packets through this MAGs BS.

For the comparison purpose, we define the relative gain of handover latency to the basic handover process as:

$$G_H^{(\cdot)} = \frac{L_H^{(\cdot)}}{L_H^{(\text{proposed})}} \quad (17)$$

Where, $G_H^{(\cdot)}$ is used for indicating a relative handover performance gain compared with other handover schemes, $L_H^{(\cdot)}$ represents the handover latency for an existing handover scheme, and $L_H^{(\text{proposed})}$ represents the proposed handover scheme. For instance, if $G_H^{(\text{intra-domain})}$ is larger than 1.0, it means that the proposed handover process outperforms the basic one.

Hence by applying equation 17 and by assuming that layer 2 time equals to 400 ms, the Gain will be:

$$G_H^{(\text{intra-domain})} = \frac{L_H^{(\text{PMIPv6})}}{L_H^{(\text{intra-domain})}}$$

$$G_H^{(\text{intra-domain})} = \frac{470}{391.5} = 1.2$$

By applying equation (17), we can notes that the gain of our novel intra-domain PMIPv6 handover technique is larger than 1.0. Hence our proposed technique outperforms the conventional PMIPv6 scheme.

IV. SYSTEM ANALYSIS RESULT

In this section, we use the parameters listed in Table 3 [1] [9] to calculate the handover latency for our proposed solution and compare it with PMIPv6 method. Figure 9 shows the

handover latency for the novel intra-domain PMIPv6. We can notice that, the proposed intra-domain PMIPv6 solution can reduce the handover latency time compared with PMIPv6. Furthermore more fig 9 shows the impact of speed on the handover process our proposed novel intra-domain PMIPv6 technique outperformed the conventional PMIPv6 scheme in reducing the overall handover processing time in vehicular network environment. Furthermore Fig 10 shows the effect of the wireless link latency impact over the handover process the results show that the novel intra-domain PMIPv6 preforms better than PMIPv6. Hence, the proposed inter-domain PMIPv6 handover process outperforms the basic handover process of PMIPv6 in terms of handover latency and data packet loss.

TABLE III. TABLE 3 PARAMETERS TO CALCULATE THE PERFORMANCE METRICS

	Notation	Default Value
Delay	T_{AAA}	30 ms
	T_{PBU}	30 ms
	T_{L2}	200-400 ms
	T_{inter}	50 ms
	T_{RA}	60
	T_{DAD}	500-1000 ms
	$T_{\text{vehicle,MAG}}$	10 ms
	b_{WD}	100 Mbps
	b_{WL}	11 Mbps
	t_{WD}	0.5 ms
	t_{WL}	2 ms
	P_f	[0.1, 0.4]
	n_h	5
	M_S^{RS}	52 byte
	M_S^{PBU}	76 bytes
	$M_S^{(HD)}$	40 bytes
M_S^{DATA}	120 bytes	
	λ_s	[0.3, 0.7]

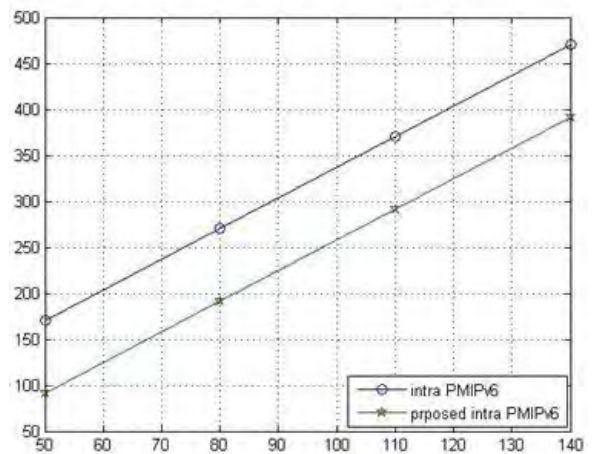


Figure 9. Handover latency vs. Vehicle Speed

V. CONCLUSION

In this paper, we proposed an intra-domain PMIPv6 handover technique for vehicular environment and compared our proposed technique with PMIPv6.

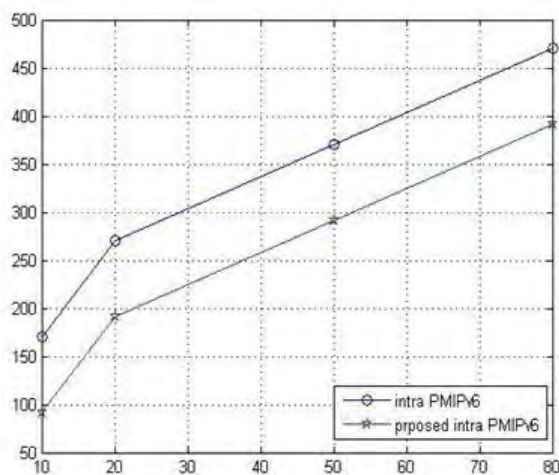


Figure 10. Handover latency vs. link latency

The proposed intra-PMIPv6 handover technique is based on MIIS information function. Using the MIH services (MIIS), the vehicle can obtain information without route discovery or RtSolPr/PrRtAd messages. Thus, the handover latency time due to concurrent start of L2 and L3 handover process is reduced. In this way, our scheme is suitable for a cost-effective network compared with PMIPv6 scheme.

The novel intra-domain handover technique is proposed to solve the high mobility of the vehicle while moving from inter-domain to intra-domain by using a logical interface to support continuance communication within the new PMIPv6 domain after completing the inter-domain PMIPv6 handover process.

Our future direction consists of plans to develop a mathematical model to evaluate the novel intra-domain PMIPv6 technique in different environments for both inter-domain and intra-domain schemes. In addition we will propose a new handover estimation technique. Further investigation of the novel intra-domain technique will be tested on a highway vehicular scenario. The impact of the intra-domain and inter-domain scheme on seamless connection support for vehicles roaming in PMIPv6 domains will be analytically analyzed.

VI. ACKNOWLEDGMENT

The authors would like to thank to the administration of Faculty of Computer Science & Information Systems and Universiti Teknologi Malaysia for their encouragement and generous support.

REFERENCES

[1] AL-Hashimi, H. N., Kamalrulnizam Abu Bakar, & Kayhan Zrar Ghafoor (2010). Inter-domain Proxy Mobile IPv6 based Vehicular Network.

Network Protocols and Algorithms.

[2] Arnold, T., W. Lloyd, J. Zhao, & G. Cao. (2008). IP address passing for VANETs. IEEE Percom (pp. 70-79). IEEE.

[3] Bechler, M., & L. Wolf. (2005). Mobility management for vehicular ad hoc networks. IEEE Vehicular Technology (pp. 2294-2298). IEEE.

[4] Fazio, M., C. P. E., S., D., & M., G. (2007). Facilitating real-time applications in VANETs through fast address auto-configuration. IEEE, 981-985.

[5] Gundavelli, S., K. Leung, V. Devarapalli, K. Chowdhury, & B. Patil. (2008, August). Proxy Mobile IPv6. Retrieved from IETF RFC 5213: <http://www.ietf.org/rfc/rfc5213.txt>

[6] Huang, C. M., M. S. Chiang, & T. H. Hsu. (2008). PFC: A packet forwarding control scheme for vehicle handover over the ITS networks. Computer Communications, 2815-2826.

[7] Johnson, D., C., P., & J, A. (2004). Mobility Support in IPv6. Retrieved from <http://www.ietf.org/rfc/rfc3775.txt>: <http://www.ietf.org/rfc/rfc3775.txt>

[8] Lee, J.-H., & Tai-Myoung Chung. (2010). Cost Analysis of IP Mobility Management Protocols for Consumer Mobile Devices. IEEE Consumer Electronics Society . IEEE.

[9] Lee, J.-H., Zhiwei Yan, & Ilsun You. (2011). Enhancing QoS of Mobile Devices by a New Handover. Springer Science+Business Media.

[10] McNair, J., Ian F. Akyildiz, & Michael B. Bender. (2000). An inter-system handoff technique for. IEEE Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM) (pp. 208-216). IEEE.

[11] Narten, & T., Nordmark, E., & Simpson W. (1998, December). Neighbor discovery for IP version 6. Retrieved from IETF RFC 2461.

[12] Sangheon Park, Ilsun You, & Tai-Myoung Chung. (2009). Enabling a paging mechanism in network-based. Journal of Internet Technology.

[13] XIAO, L., & Jiawei YANG. (2010, 08 20). Performance analysis of proxy mobile IPv6 based on IEEE802.16e.

[14] Yan, Z., Huachun Zhou, & Ilsun You. (2010). N-NEMO: A comprehensive network mobility solution in proxy mobile IPv6 network. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 52-70.

AUTHORS PROFILE

Haidar N. AL-Hashimi is currently a PhD candidate studying in the Department of Computer Network & Communication, Universiti Teknologi Malaysia, Johor, Malaysia. Haidar received his B.Sc degree in Computer Science and M.Sc degree in Computer Science/Unicast Routing in Mobile Ad hoc Networks in 2005 and 2009 respectively. His current research interests include Vehicular Ad Hoc Networks, Routing over VANET, and Handover Process in Heterogeneous Vehicular Networks.

Kamalrulnizam Abu Bakar obtained his PhD degree from Aston University (Birmingham, UK) in 2004. Currently, he is associate professor in Computer Science at Universiti Teknologi Malaysia (Malaysia) and member of the "Pervasive Computing" research group. He involves in several research projects and is the referee for many scientific journals and conferences. His specialization includes mobile and wireless computing, information security and grid computing.

Shaharuddin Salleh obtained his PhD degree from Universiti Teknologi Malaysia, Johor, Malaysia in Computational Mathematics (UTM) Currently, he is professor in Mathematic Department at Universiti Teknologi Malaysia (Malaysia). He involves in several research projects and is the referee for many scientific journals and conferences. His specialization Computational Mathematics & Network modeling.

Autonomous Control of Eye Based Electric Wheel Chair with Obstacle Avoidance and Shortest Path Findings Based on Dijkstra Algorithm

Kohei Arai

Graduate School of Science and Engineering
Saga University
Saga City, Japan

Ronny Mardiyanto

Graduate School of Science and Engineering
Saga University
Saga City, Japan

Abstract—Autonomous Eye Based Electric Wheel Chair: EBEWC control system which allows handicap person (user) to control their EWC with their eyes only is proposed. Using EBEWC, user can move to anywhere they want on a same floor in a hospital autonomously with obstacle avoidance with visible camera and ultrasonic sensor. User also can control EBEWC by their eyes. The most appropriate route has to be determined with avoiding obstacles and then autonomous real time control has to be done. Such these processing time and autonomous obstacle avoidance together with the most appropriate route determination are important for the proposed EBEWC. All the required performances are evaluated and validated. Obstacles can be avoided using acquired images with forward looking camera. The proposed EBEWC system allows creation of floor layout map that contains obstacles locations in a real time basis. The created and updated maps can be share by the electric wheel chairs on a same floor of a hospital. Experimental data show that the system allows computer input (more than 80 keys) almost perfectly and electric wheel chair can be controlled with human eyes-only safely.

Keywords- Human Computer Interaction; Gaze Estimation; Obstacle Avoidance; Electric Wheel Chair Control.

I. INTRODUCTION

Electric Wheel Chair: EWC for mobility support for handicap persons, in particular, is presented [1]-[5]. There are some the proposed systems for mobility support using eye movements based on Electrooculography. EWC is controlled by eye movement which is acquired using Electrooculograph [1]. Also there is the proposed integrated solution to motion planning and control with three input sources [2]. User can specify their designated destination using a visual interface by looking at the designated functional keys displayed onto computer screen. This input makes the EWC control automatically and is generated by a deliberative plan incorporating with a prior knowledge. User also can use reactive controller to avoid obstacles and features that the sensor detect. User can directly provide velocity command using joystick or some other input devices as a basic level. There is the proposed vision-based navigation system for an electric wheelchair using ceiling light landmark [3]. The EWC is equipped with two cameras those are used for self location and obstacle avoidance. The fluorescent ceiling lights are

chosen as landmarks since they can be easily detected and do not require an additional installation. Also there is the proposed head gesture based control of an intelligent wheelchair [4]. This system used Adaboost face detection¹ which is provided by OpenCV. By detecting frontal face and nose position, head pose and gesture are estimated and are used for control the EWC. There is the proposed EWC control by the detection of gaze direction and eye blinking [5]. The gaze direction is expressed by horizontal angle of gaze. It is derived from the triangle form which is formed by the center position of eyes and nose. The gaze direction and eye blinking are used to provide the direction and timing command. The direction command related to the move direction of EWC and the timing command related to the time condition when EWC has to be moved.

In order to estimate gaze based on image analysis, it is common that gaze location is estimated with pupil location. The previous pupil detection methods can be divided into two categories: the active Infrared: IR-based approaches [7], [8], [16] and the traditional image-based passive approaches [9]-[15]. Eye detection based on Hough transform is proposed [9]-[11]. Hough transform is used for finding the pupil. Eye detection based on motion analysis is proposed [6], [7]. Infrared light is used to capture the physiological properties of eyes (physical properties of pupils along with their dynamics and appearance to extract regions with eyes). Time series analysis of motion such as Kalman filter² and mean shift which are combined with Support Vector Machine: SVM³ used to estimate pupil location. Eye detection using adaptive threshold and morphologic filter⁴ is proposed [12]. Morphologic filter is used to eliminate undesired candidates for an eye and pupil. Hybrid eye detection using combination between color, edge and illumination is proposed [15].

We proposed the computer input system with human eye-only and it application for EWC control [17]. It is called Eye Based EWC: EBEWC hereafter. EBEWC works based on eye gaze. When user looks at appropriate angle/key, then computer

¹ <http://note.sonots.com/SciSoftware/haartraining.html>

² http://en.wikipedia.org/wiki/Kalman_filter

³ <http://www.support-vector.net/>

⁴ <http://www.spm.genebee.msu.ru/manual/en/node108.html>

input system will send command to EWC. EBEWC is controlled by three keys: left (Turn left), right (Turn right), and down (Move forward) in order to give command to electric wheelchair: turn left, turn right, and go forward [18]. This three combination keys are more safely than others combination keys. The stop key is not required because EWC automatically stops when user changes the gaze. Gaze estimation performance is well reported in the previous paper [19] in terms of robustness against a variety of user types, illumination changes, vibration as well as gaze position accuracy, and calibration accuracy. Also EBEWC with obstacle avoidance control is described in the previous paper [20]. In this paper, automatic EWC path finding algorithm and EWC control performance are discussed together with performance to ensure a real-time processing of all the required processes for autonomous control, in particular, obstacle avoidance with ultrasonic sensors and visible cameras.

II. PROPOSED EYE BASED ELECTRIC WHEEL CHAIR: EBEWC CONTROL

A. Hardware Configuration

Fig.1 shows the hardware configuration of the proposed EBEWC. The proposed system consists of (1) two cameras mounted glass, (2) NetBook PC, (3) Ultrasonic Sensor. Two cameras look at forward (finding obstacles) and backward (acquires users' eye image). The camera used 1.3 Mega pixel OrbiCam (Visible camera) and IR camera for gaze estimation.

The Yamaha JW-I type⁵ of EWC is used. Netbook of Asus EeePC with 1.6 GHz Intel Atom processor, 1GB of RAM, 160GB hard drive, and run Windows XP Home edition is also used. We develop our software under C++ Visual Studio 2005 and Intel provided OpenCv image processing library [6]. The proposed EBEWC system utilizes infrared web camera, NetCowBoy DC-NCR 131 as face image acquisition which works in a real time basis. This camera has IR Light Emission Diode: LED. Therefore, it is robust against illumination changes. Furthermore, pupil detection becomes much easy. In order to allow user movement and EWC vibration, IR camera mounted glass is used. The distance between camera and eye is set at 15.5 cm. The Yamaha JW-1 type of EWC is controlled by human eyes only through microcontroller of AT89S51⁶. This microcontroller can convert serial output from the Netbook to digital output for control.

B. EWC Control by Human Eyes Only

In order to control EWC, at least four keys, move forward, turn right, turn left and stop are required. For the safety reason, users have to look forward so that the key layout of Fig.2 is proposed. Namely, key consists of 9 keys (3 by 3). Move forward and turn left/right are aligned on the middle row. Stop key is aligned on the other rows, top and bottom rows. Users understand the location of desired key so that it can be selected with users' eye-only without any computer screen. Basic actions of the EBEWC control are as follows,

- If eye detection fail then stop
- If obstacle detected then stop
- If user surprise then stop

for safety reason. When user surprises, then typical eye shape becomes open widely and usually upward looking shape.

The backward looking camera whose resolution is 640 by 480 pixels acquires users' eye and its surrounding. Using OpenCV of eye detection and tracking installed on the Netbook PC, users' eye is detected and tracked. If the OpenCV cannot detect users' eye, then EWC is stopped for safety reason. EWC is also stopped when users look at the different location other than the three keys aligned on the center row.

Intentional blink can be detected if the eye is closed for more than 0.4 seconds because accidental blink is done within 0.3 seconds, typically. In this connection, it is easy to distinguish between intentional and accidental blink. Also, key selection can be done every 0.4 seconds. Thus the system recognizes user specified key every 0.4 seconds. In order to make sure the user specified key, 10 frames per seconds of frame rate is selected for backward looking camera.

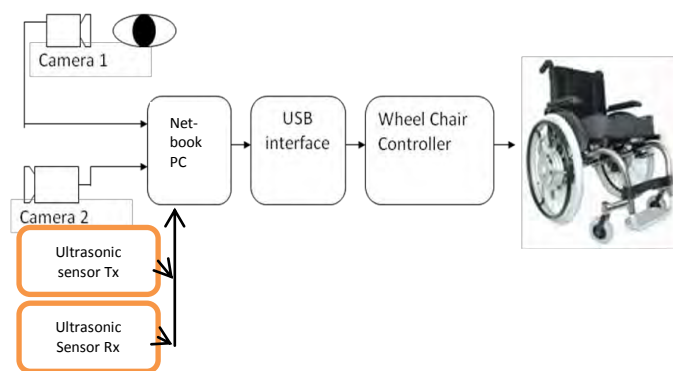


Figure 1 Hardware configuration of the proposed EBEWC system

Switch	Stop	Stop
Turn left	Forward	Turn right
Stop	Stop	Stop

Figure 2. 3 by 3 of key layout for EWC control

C. Autonomous EWC Guide to Destinations

If user selects the top left corner of key, then the key layout is changed to Fig.3. After that, user can select previously designated destination, 8 candidate destinations in this case. User selects the destination previously. For instance, nurse station is destination #1, rest room is destination #2 and so on.

Switch	Destination #1	Destination #2
Destination #3	Destination #4	Destination #5
Destination #6	Destination #7	Destination #8

Figure 3. 3 by 3 of key layout for EWC control

In this case, EBEWC goes to the destinations autonomously with a reference to the floor layout and with avoiding obstacles. Obstacle avoidance can be done with ultrasonic sensor and forward looking visible camera. The ultrasonic sensor is for transparent obstacle such as glass door, window, etc. and detection as well as avoidance in particular.

⁵ <http://disabled-help.webeden.co.uk/#/yamaha-jw1/4514819994>

⁶ http://www.atmel.com/dyn/resources/prod_documents/doc2487.pdf

When EWC detect the obstacle, it must understand where the best path should be chosen if EWC want to go to specific place. The most appropriate route will be chosen based on floor layout map and image map. Image map is created by acquiring background images in every 1 m as is shown in Fig.4. Medical doctors, patients and nurses are obstacles. Even if the location cannot be detected with forward looking visible camera and ultrasonic sensor, their location can be acquired with wireless LAN connections. Access points of antennas are equipped at the four corners of a floor in a hospital. Therefore, the locations can be determined by using the delay times from the four antennas. These location information are gathered and are provided to the all the EWC on a floor. Thus EWC can avoid obstacles.

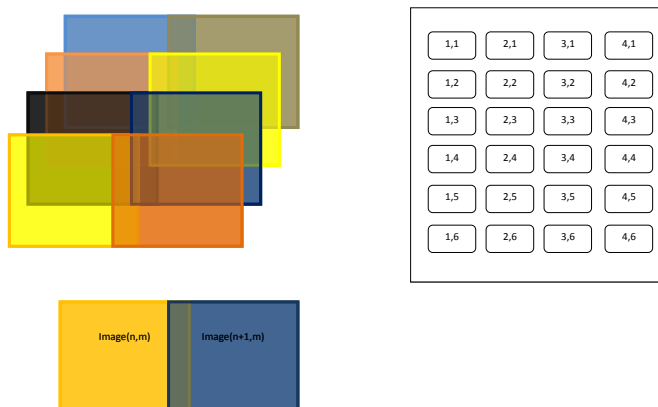


Figure 4 Image map, created by acquire image in every 1 m.

Before getting start to operate EBEWC, a template of one shot image is acquired with the forward looking camera at each location (x, y) . For instance, if area of a floor in a hospital is 10 m by 10 m, then it takes 100 images. As an initial condition, a floor layout without any obstacles of pedestrians, medical doctors, nurses, patients, etc. is set-up. After acquiring background of template images, a floor layout map with obstacles is updated automatically every unit time. This map is setup manually based on room layout. Example is shown in Fig.5.

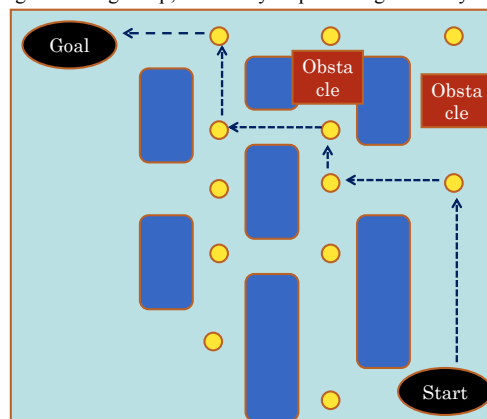


Figure 5 Shortest path finding using Dijkstra

Combination between background images and layout path will obtain main map. EWC will move consider the main map. After obstacle is detected in current path, EWC will switch to another path base on Dijkstra algorithm⁷ with the same destination. Obstacle avoidance methods are also useful when user is not confident to pass through the obstacle. EWC will avoid the obstacle. Furthermore, the proposed EBEWC system also is able to update existing map. In this example, a floor is to be 4 by 6 m. Image is acquired at each lattice point location with 1m of intervals and with 10% of overlapping between the neighboring two areas. Then forward looking camera acquires same location of image. After that, the current image is compared to the previously acquired image results in obstacle finding. Namely, the different between the current and the previously acquired images shows location and shape of obstacles. EBEWC may switch back to the original key layout when user looks at the top left corner of key for much longer than 0.7 sec.

D. Obstacle Avoidance

For safety reason, obstacle avoidance system is implemented in the proposed EBEWC system. Obstacle avoidance system is capable to identify the obstacle in front of EWC and avoid it. This system consists of two approaches: (1) Obstacle detection, and (2) Shortest Path Finding. Obstacle detection is consisting of image processing based and ultrasonic sensor based. Image processing based utilizes background subtraction between current image $I(x,y)$ and background image $B(x,y)$. Background subtraction method will obtain black-white image $S(x,y)$ which represent obstacle. On this image, obstacles appear as white pixel. By using searching of outer line from white pixels, we can determine position and size of obstacle.

$$S(x, y) = 0, \text{ if } |I(x, y) - B(x, y)| < \text{threshold} \quad (1)$$

$$S(x, y) = 255, \text{ if } |I(x, y) - B(x, y)| \leq \text{threshold} \quad (2)$$

Weakness of background subtraction method is working only if two images have same position of translation, scale, shear, and rotation. To solve these problems, we utilize Affine transformation⁸. This transformation requires three points that appears on both images.

Translation, scale, share, and rotation parameter can be determined from these points. Affine transformation required at the least three noticeable important points. These points should be appears on both images. It can be detected by several ways: corner, edge, specific object, and etc. Corner and edge have many points and its will create computation problem. In our system, we decide to use specific object. The specific object can be an easily recognized object, text character, chessboard wall, and etc.

The specific object will obtain one coordinate from center of area. Obstacle detection using Affine transformation requires identifying at least three kind of object, so it will obtain three noticeable important points. First step of obstacle detection is convert source image into gray image. By using template matching, system will find specific object position. Using six noticeable points, system calculate translation, share, scale, and

⁷ <http://mathworld.wolfram.com/DijkstrasAlgorithm.html>

⁸ <http://mathworld.wolfram.com/AffineTransformation.html>

rotation parameter. These parameters will be used for creating Affine transformation matrix.

Affine transformation matrix is applied on background image will obtain Affine transformed image. Normalization is used to eliminate disturbance such shadow, noise, and etc. Next, subtract between output from current image and background image. Smooth filter is used to reduce noise which is caused by subtraction process. Last step is applying threshold on image and it will obtain black and white image. Obstacle will be signed as group of white pixel. After black white image is obtained, we should return center of white pixel area into current coordinate and coordinate of obstacle is founded. Because of so many type of specific object, we got best performance of specific object by using chessboard wall. The advantage of the chessboard (Fig.6) is easily detected and robust on distance changes. We use three types of chessboard: 3 by 4, 3 by 5, and 3 by 6 as are shown in Fig.7. The other obstacle detection is use ultrasonic sensor. This sensor has advantage when visual system does not work. In case EWC move into surrounding glass door, smoke condition, and minimum lighting will caused visual system of obstacle detection fail. Ultrasonic sensor consists of transmitter and receiver part. We use ultrasonic sensor with 40 kHz frequency resonance.

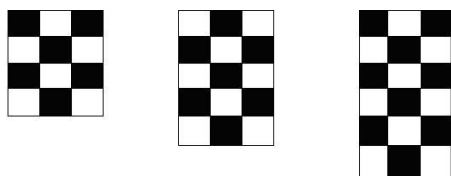


Figure 6 Chessboard as specific point

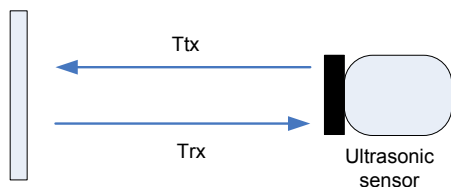


Figure 7 Ultrasonic sensor, distance is measured by converting half of Tx transmitting time and Trx receiving time.

E. Shortest Path Findings

EWC will move consider the main map. After obstacle is detected in current path, EWC will switch to another path which have same destination. Obstacle avoidance methods are also useful when user is not confident to pass the obstacle. EWC will take over and pass the obstacle. Best path is chosen based on Dijkstra algorithm. Furthermore, system also able to renew existing map.

III. EXPERIMENTS

A. Obstacle Avoidance Performance

Experiment of obstacle avoidance is conducted by acquiring image in corridor with distance 1m per images. Location which image was acquired, is set with (x,y) coordinate. So, in every (x,y) location will have 1 background image. This image will be used as background reference. Obstacle is detected by subtract background image with current image. To

eliminate problem which caused by different position between these images, Affine transformation is used to transform background image to Affine transformed image which have same position with current image. Applying Affine transformation will does work if only if three noticeable important points are appears on both images. These important points are represented by three types of chessboard (Fig.6). If the chessboard is successful detected, then by using Affine transformation and subtraction between background image and current image, obstacle is founded. Experiments of chessboard detection is conducted by measuring maximum location which still detected by system. Chessboard is put in fix location. After that, EWC is move around the chessboard. Maximum location which detected by system is recorded. Data show that boundary shift will decrease when distance between camera and chessboard is increase. This experiment is equal to obstacle detection. After three types of chessboard are detected, Affine transformation will use these points (three chessboard center of areas) to subtract with current image and obstacle position will be found.

Objective of this experiment is measure accuracy of ultrasonic sensor before used in the system. This experiment is conducted by measure sensor output for measuring distance and comparing with real data. Some object is put on front of sensor with varies distances and measure it. The experiment is conducted on distance 0 cm until 3 m. Ultrasonic sensor use PING type parallax product⁹ and microcontroller AT89S51 as processing data (convert from time value to distance output). PING type parallax product and microcontroller AT89S51 as processing data (convert from time value to distance output). Elevation angle is require to know how width of the beam sensor. Ultrasonic sensor with width beam is not benefit to our system. Narrow beam will obtain good result because it will not influence with any disturbance. This experiment is conducted by measure elevation angle from 0 cm until 3 m.

B. EWC Control Performance

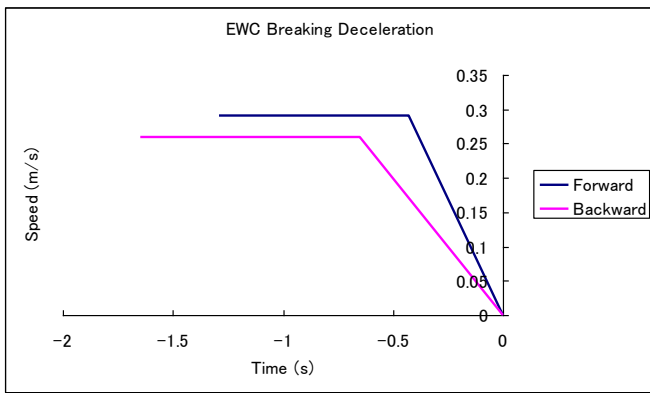
This experiment is conducted for measure EWC performance on starting up acceleration, forward and backward breaking deceleration. Also conducted speed measurement when EWC move forward, backward, turn left, and turn right. EWC is drive by user who has weight is 73 kg. We record the duration time and convert it into speed value. Experiment data of speed measurement is shown in Table 1. Graph of EWC acceleration and deceleration when start and stop duration is shown in Fig.8.

C. Processing Time for Each Process

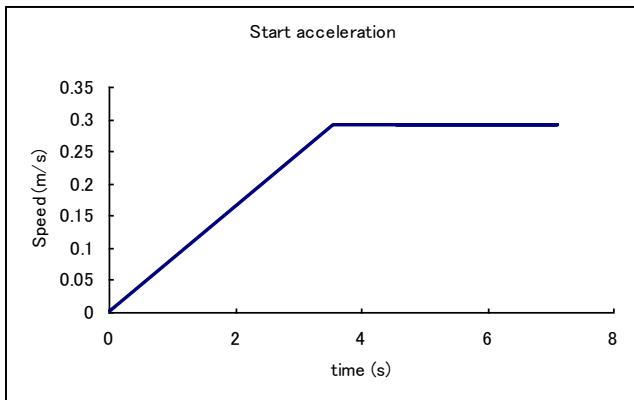
In order to apply whole method into EWC application, processing time should be measured to identify performance of our real time system. Fig.9 shows transient time of eye detection and tracking, n the beginning of chart, it seem this method take long time around 300 ms. In this time, system still process face detection, eye detection and others process before running template matching method.

⁹

<http://www.parallax.com/tabid/768/ProductID/92/Default.aspx>



(a)Deceleration



(b)Acceleration

Figure 8 Experiment of EWC acceleration and deceleration

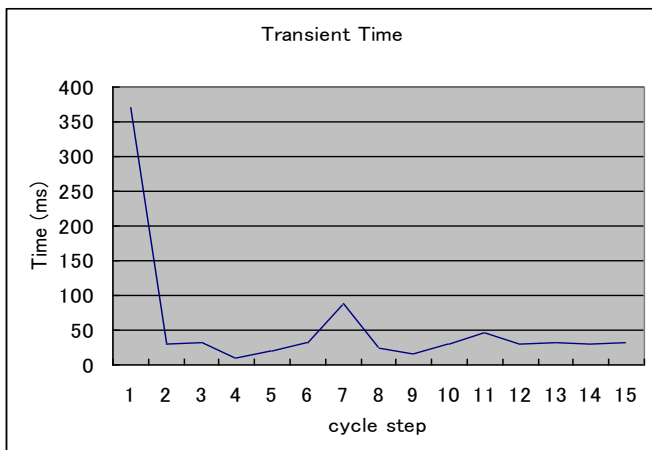


Figure 9 Transient time of eye of detection and tracking, n the beginning of chart, it seem this method take long time around 300 ms. In this time, system still process face detection, eye detection and others process before running template matching method. After eye location is founded, then system bypass previous step and cause process working fast.

After eye location is founded, then system bypass previous step and cause process working fast. Meanwhile, Fig.10 shows processing time of eye detection and tracking on steady state condition, it looks faster than transient condition.

This experiment is conducted using Optiplex 755 Dell

computer with Intel Core 2 Quad CPU 2.66 GHz and 2G of RAM. We use NET COWBOY DC-NCR131 camera as visual input. Experimental result show average steady state processing time is 32.625 ms. it also shows difference processing time between transient and steady state condition. Transient time require more time than steady state time.

Objective of this experiment is measure processing time on Eye gaze identification. It is conducted by using ACER computer ASPIRE 5572 Series Core Duo T2050 1.6 GHz CPU and 1G of RAM. Result data show average processing time of this method is 342.379 ms. Fig.11 shows processing time of Eye gaze method.

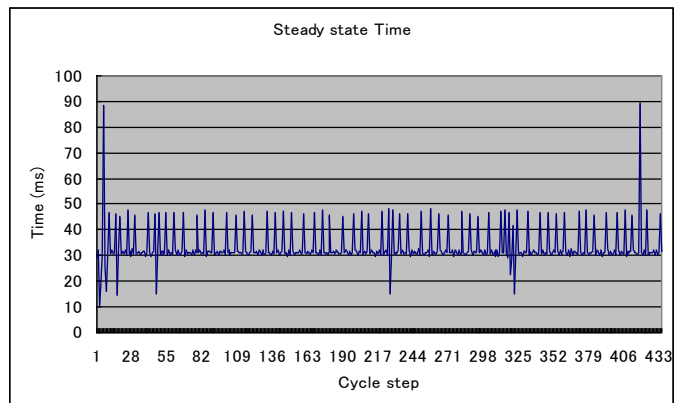


Figure 10 Processing time of eye detection and tracking on steady state condition, it looks faster than transient condition.

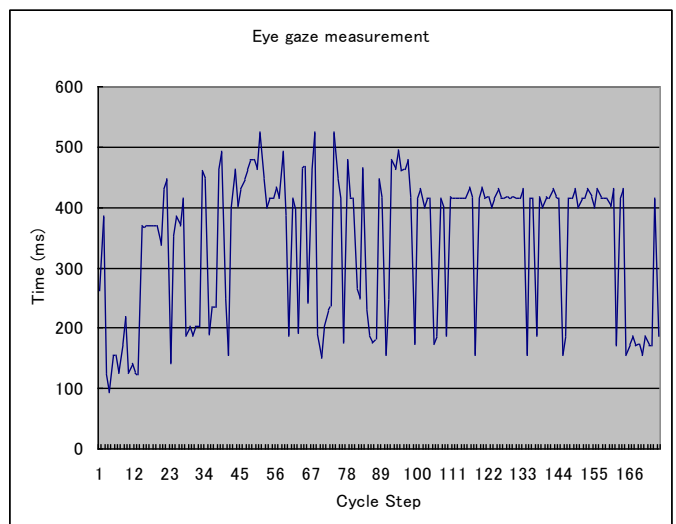


Figure 11 Processing time of Eye gaze method

TABLE.1 SPEED MEASUREMENT

Moving	Speed (m/s)
Forward	0.29
Backward	0.26
Turn Left	0.11
Turn Right	0.11

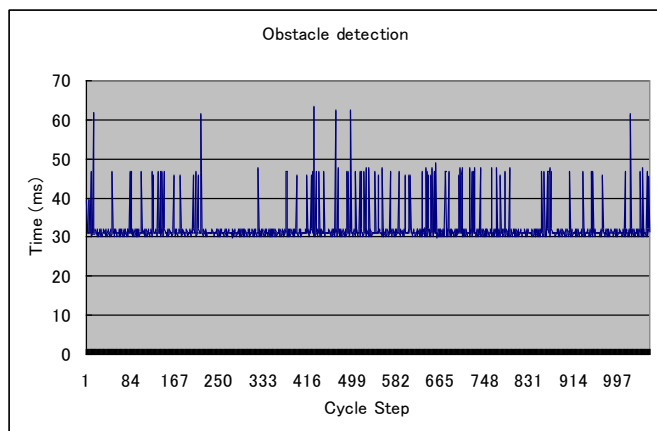


Figure 12 Processing time of obstacle detection

This experiment also was conducted using Optiplex 755 Dell computer with Intel Core 2 Quad CPU 2.66 GHz and 2G of RAM. NET COWBOY DC-NCR131 camera as visual input is also used. Experimental result show average processing time is 32.625 ms. Fig.12 shows processing time of obstacle detection.

We implemented ultrasonic sensor parallax PING type. This sensor is controlled by using custom microcontroller AT89S51. Data was stored into computer by using USB communication. Result data show average processing time is 568.658 ms. Fig.13 also shows processing time of ultrasonic sensor, it look take longer time than others.

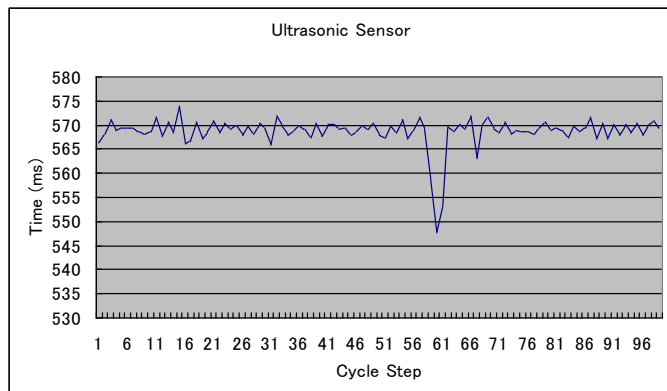


Figure 13 processing time of ultrasonic sensor, it look take longer time than others.

IV. CONCLUSIONS

It is concluded that the proposed key-in system with human eyes only works 100% perfectly for the number of keys is four, start, stop, turn right and left. Also it is concluded that the proposed EWC control system does work in a real basis avoiding obstacles on range bellow 3.4 m using image processing method and bellow 3 m using ultrasonic sensor. By the proposed system, EWC is able to identify obstacle and avoid them. Obstacle avoidance can assist user into interest place although undesired condition such as dark areas, glass wall or door, smoke area, and etc. By implemented this system, it is found that EWC can move more safely and comfortable.

We may conclude the followings,

- 1) EWC will stop after 70cm when load below than 73 Kg
- 2) EWC will stops after 30 cm when 85 Kg
- 3) Increasing load will make EWC stop shorter
- 4) Proposed system is almost perfect for detection of functional keys which selected by human eye only.
- 5) System is capable to detect the obstacle on range bellow 3.4 m.

Also it is confirmed that all the required processing can be done in a real time basis.

ACKNOWLEDGMENT

Authors would like to thank to the graduate students who contributes to the performance evaluation experiments of the proposed EBWC.

REFERENCES

- [1] Barea, R., Boquete, L., Mazo, M., Lopez, E.: "System for Assisted Mobility using Eye Movements based on Electrooculography", IEEE Transaction on Neural System and Rehabilitation Engineering, 10, 4, 209-218, 2002.
- [2] Sarangi, P., Grassi, V., Kumar, V., Okamoto, J.: "Integrating Human Input with autonomous behaviors on an Intelligent Wheelchair Platform", Journal of IEEE Intelligent System, 22, 2, 33-41, 2007.
- [3] Wang, H., Ishimatsu, T.: "Vision-based Navigation for an Electric Wheelchair Using Ceiling Light Landmark", Journal of Intelligent and Robotic Systems, 41, 4, 283-314, 2005.
- [4] P. Jia and H. Hu: "Head Gesture based control of an Intelligent Wheelchair", Proceedings of the 11th Annual Conference of the Chinese Automation and Computing Society in the UK [CACUSUK05], 85-90, 2005.
- [5] D. Purwanto, R. Mardiyanto, K. Arai: "Electric wheelchair control with gaze direction and eye blinking", Proceedings of The Fourteenth International Symposium on Artificial Life and Robotics, GS21-5, B-Con Plaza, Beppu, 2008.
- [6] Gary Bradski, Andrian Kaehler: "Learning Computer Vision with the OpenCV Library", O'REILLY, 214-219, 2008.
- [7] Haro, A, Flickner, M, Essa, I: "Detecting and Tracking Eyes By Using Their Physiological Properties, Dynamics, and Appearance ", Proceedings of the CVPR 2000, 163-168, 2000
- [8] Zhiwei Zhu, Qiang Ji, Fujimura, K., Kuangchih Lee: "Combining Kalman filtering and mean shift for real time eye tracking under active IR illumination", Proceedings of the 16th Pattern Recognition International Conference, 4, 318- 321, 2002.
- [9] Takegami, T, Gotoh, T, Kagei, S, Minamikawa-Tachino, R: "A Hough Based Eye Direction Detection Algorithm without On-site Calibration", Proceedings of the 7th Digital Image Computing: Techniques and Applications, 459-468, 2003.
- [10] K.M.Lam, H.Yan, Locating and extracting eye in human face images, Pattern Recognition, 29, 5, 771-779, 1996.
- [11] G.Chow, X.Li, Towards a system of automatic facial feature detection, Pattern Recognition 26, 1739-1755, 1993.
- [12] Rajpathaka, T, Kumarb, R, Schwartzb, E: Eye Detection Using Morphological and Color Image Processing", Proceedings of the Florida Conference on Recent Advances in Robotics, 2009.
- [13] R.Brunelli, T.Poggio, Face Recognition, Features versus templates, IEEE Trans. Patt. Anal. Mach. Intell. 15, 10, 1042-1052, 1993.
- [14] D.J.Beymer, Face Recognition under varying pose, IEEE Proceedings of the Int. Conference on Computer Vision and Pattern Recognition [CVPR'94], Seattle, Washington, USA, 756- 761, 1994
- [15] Shafi, M, Chung, P. W. H: "A Hybrid Method for Eyes Detection in Facial Images", International Journal of Electrical, Computer, and Systems Engineering, 231-236, 2009.
- [16] Morimoto, C., Koons, D., Amir, A., Flickner, M: "Pupil detection and tracking using multiple light sources", Image and Vision Computing, 18,4, 331-335, 2000
- [17] Kohei Arai, Hiromi Uwataki, Computer input system by human eyes only based on line of sight estimation allowing users' movements,

Journal of Institute of Electric Engineering of Japan, C-127, 7, 1107-1114, 2007.

- [18] Djoko Purwanto, Ronny Mardiyanto and Kohei Arai, Electric wheel chair control with gaze detection and eye blinking, *Artificial Life and Robotics, AROB Journal*, 14, 694,397-400, 2009.
- [19] Kohei Arai and Makoto Yamaura, Computer input with human eyes only using two Purkinje images which works in a real time basis without calibration, *International Journal of Human Computer Interaction*, 1,3, 71-82,2010
- [20] Kohei Arai, Ronny Mardiyanto, A prototype of electric wheel chair

control by eye only for paralyzed user, *Journal of Robotics and Mechatronics*, 23, 1, 66-75, 2010.

AUTHORS PROFILE

Kohei Arai received a PhD from Nihon University in 1982. He was subsequently appointed to the University of Tokyo, CCRS, and the Japan Aerospace Exploration Agency. He was appointed professor at Saga University in 1990. He is also an adjunct professor at the University of Arizona since 1998 and is Vice Chairman of ICSU/COSPAR Commission A since 2008

Eye-based Human Computer Interaction Allowing Phoning, Reading E-Book/E-Comic/E-Learning, Internet Browsing, and TV Information Extraction

Kohei Arai

Dept. of Information Science
Saga University
Saga City, Japan

Ronny Mardiyanto

Dept. of Information Science
Saga University
Saga City, Japan

Abstract—Eye-based Human-Computer Interaction: HCI system which allows phoning, reading e-book/e-comic/e-learning, internet browsing, and TV information extraction is proposed for handicap student in E-Learning Application. The conventional eye-based HCI applications are facing problems on accuracy and process speed. We develop new interfaces for improving key-in accuracy and process speed of eye-based key-in for E-Learning application, in particular. We propose eye-based HCI by utilizing camera mounted glasses for gaze estimation. We use the sight for controlling the user interface such as navigation of e-comic/e-book/e-learning contents, phoning, internet browsing, and TV information extraction. We develop interfaces including standard interface navigator with five keys, single line of moving keyboard, and multi line of moving keyboard in order to allow the aforementioned functions without burdening the accuracy. The experimental results show the proposed system does work the aforementioned functions in a real time basis.

Keywords-Eye-based HCI; E-Learning; Interface; Keyboard.

I. INTRODUCTION

Recently, the development of eye-based human computer interaction is growing rapidly. This grow is influenced by the growing of the number paraplegics. The number of paraplegics extremely increased (It was reported that in 2009 the number of paraplegics in U.S.A. has gained up 40% from 2007) caused by accident working (28%), motor vehicle accident (24%), sporting accident (16%), fall (9%), victim of violence (4%), birth defect (3%), natural disaster (1%), and others [1].

Nowadays, the eye-based Human-Computer Interaction: HCI has been widely used to assist not only handicap person but also for normal person. In handicap person, especially paraplegic, they use eye-based HCI for helping them to self-sufficient in the daily life such as input text to computer [2], communication aids [3], controlling wheelchair [4] [5], having meal on table using robot arm [6], etc. The eye key-in system has been developed by many researchers [2]. The commercial available system provided by Tobii Tracker Company¹ has been used by many researchers for developing text input, customer interest estimator on business market, etc [7].

Technology has been successful at rehabilitating paraplegics' personal lives. Prof. Stephen Hawking², who was diagnosed with Amyotrophic lateral sclerosis³ (ALS), uses an electronic voice synthesizer to help him communicate with others [8]. By typing the text through aid of a predictive text entry system, approximating his voice, he is able to make coherent speech and present at conferences. To give another example, a paraplegic patient wearing a head-mounted camera is able to draw figures, lines, and play computer games [2]. Clearly, through use of assistive technology, handicapped people are able to do feats on par with non-handicapped people.

The published papers discussing eye-based HCI system are categorized into: (1) vision-based and (2) bio-potential-based. The vision-based method utilized camera to capture image and estimate the user sight. The key issue here is how the method/system could be deal with environment changing. Lighting changing, user movement, various types of user, etc have to cooperate with the system. The vision-based system could be explained as follows,

- 1) Ref. [9] developed eye mouse based on user's gaze. After face is found and tracked, eye location is searched by projection of difference between left and right eye images. Output of this system is only left and right direction which used to control mouse pointer. No upward and downward directions are used. It has been implemented to control application such "BlockEscape"⁴ game and spelling program.
- 2) Ref. [10] developed eye mouse which user's gaze is obtained from pupil location by using Haar Classifier (OpenCv function⁵). Also, blinking is used as left click mouse event.
- 3) Ref. [11] developed camera mouse using face detection and eye blink. Center position of face is detected by using Adaboost⁶ face detection method and tracked by using

² http://en.wikipedia.org/wiki/Stephen_Hawking

³ http://en.wikipedia.org/wiki/Amyotrophic_lateral_sclerosis

⁴ BlockEscape

⁵ <http://sourceforge.net/projects/opencvlibrary/>

⁶ <http://note.sonots.com/SciSoftware/haartraining.html>

¹ <http://www.tobii.com/>

Lucas-Kanade method⁷. This location is used as pointing value and blinking is used as left click mouse event.

- 4) Ref. [12] developed a human-computer interface by integrating eye and head position monitoring devices. The system was controlled based on user sight and blinking. The user command could be translated by system via sight and blinking. Also, they modified calibration method for reducing visual angle between center of target and the intersection point (derived by sight). It was reported that this modification could allowed 108 or more command blocks to be displayed on 14 inch monitor. Also, it was reported that it has hit rate of 98% when viewed at the distance of 500mm. For triggering, the blinking was used to invoke commands.

The bio-potential-based method estimated user behavior (eye behavior) by measuring user's bio-potential. The bio-potential measurement instrument is required for measuring eye behaviors. The example of bio-potential-based has been applied into application of electric wheelchair controlled using Electrooculograph (EOG)⁸ analyzed user eye movement via electrodes directly on the eye to obtain horizontal and vertical eye-muscle activity. Signal recognition analyzed Omni directional eye movement patterns [13].

In this paper, we propose eye-based HCI allowing phoning, reading E-Book, E-Leaning and E-Comic, and TV information extraction. E-Book, E-Learning, E-Comic contents can be accessible through Internet. The proposed Eye Based Tablet PC: EBTPC allows read the contents [14]. One segmentation TV signal can be acquired with tuner. Sometime users would like to get information for purchasing products introduced from the TV program. The conventional system need human resources to extract such information for purchasing the products then create sales product database for consumers. They used to sell access fees for getting information. The proposed TV information extraction allows users to extract information automatically [15]. The objective of this research is how we could use to replace the use of touch screen that always rely on hand. The use of touch screen to input a command has been widely used in many applications. The use of it is still limited only for normal person who could input a command by touching the screen directly using hand. Unfortunately, the handicap person will not be able to use it since he could not use his hand to input a command via touch screen like the normal person. In this research, besides allowing the use of it for handicap person, it should improve the response time of typing since the sight is faster than hand control. If we input a command using hand, it could be fast if hand have recognized the location of the key, unfortunately the actual speed rely on the distance between key. If the bigger size of keyboard is used, the hands-typing speed will decrease (It happens if the distance among the keys is farther than the finger covered area). For the condition with distance among keys is farther, the sight will be faster than the hands. In this research, we propose eye-based HCI by utilizing camera mounted on user glasses to estimate the user sight. By fixing the users head position, we estimate the sight of user to

display. We use the sight detection result to input command such as navigate E-Comic/E-Book/E-Learning reader, call a phone number for phoning, and TV information extraction.

II. PROPOSED METHOD

In this paper, we propose a new system of eye-based HCI allowing phoning, browsing internet, reading E-book/E-Comic, and TV information extraction. Such system will help handicap person using mobile E-Learning system.

The mobile E-Learning system utilizing mobile phone or smart phone for accessing E-Learning content from server have to be prepared also for handicap student. The handicap student who has difficulty to use hands will face problems using E-Learning system. Beside the application for E-Learning system, we also prepare it for making call (allowing user type phone number and making phone call), internet browsing, E-Comic/E-Book/E-Learning content reader, and TV information extraction. In this system, we design user interface that will be explained as follows,

A. Proposed User Interface

Error! Reference source not found. shows the interface's flow of our system. The startup menu will show buttons consisting of Phone Dial Pad, E-book/E-Comic Reader, Internet browsing, and TV Information extraction. These buttons provide different functionality that could be choosing by user. User chooses one of this menu buttons to enter the sub menu. The main menu is shown in **Error! Reference source not found.**

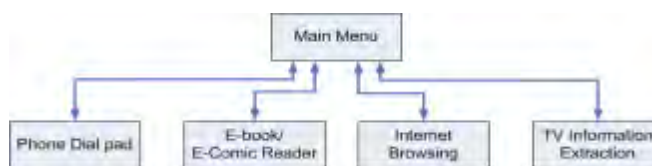


Figure 1 User Interface



Figure 2 Main menu



Figure 3 Phone dial pad sub menu

It consists of four buttons placed in top (Phone dial pad), down (TV), right side (Internet Browsing), and left sides (Read E-Comic/E-Book). To select the button, user have to look at it and hold within few seconds (using timer mode) or

⁷ http://en.wikipedia.org/wiki/Lucas%E2%80%93Kanade_method

⁸ <http://en.wikipedia.org/wiki/Electrooculography>

blinking (using blinking mode) for execution. In our system, we maintain the number of button is five (top, down, left, right, and center) for maintaining selection accuracy (As we know that the increasing the number of key/button will decrease the accuracy). Also, we design these buttons with same distance among them for making all buttons become a button with same characteristic with others.

After user select the button on main menu, the sub menu of selected button will be shown. If user selects the phone dial pad button, the interface such in **Error! Reference source not found.** will appears.

It contains four buttons with single line moving keyboard. The four buttons are used to move the moving keyboard to left or right, call the selected phone number, and return to main menu via “BACK” button. The single line moving keyboard consist of the number characters and symbol that as used in usual phone dial pad. We only use single line moving keyboard because the number of character for phone dial pad is few, so it does not need multi line moving keyboard. User could select the phone number by navigating the left and right button to move the single line moving keyboard. User have to locate the candidate of selected number to center by using these two navigator buttons (“LEFT” and “RIGHT”). To locate the candidate of selected number to center (for instance “4”), user could look at “LEFT” within 2 steps (if the initial condition is “6”) until the “4” moves to center (the number located in center will be shown in bigger size to help user distinguish it easily).

The other sub menu is E-book/E-Comic reader as shown in **Error! Reference source not found.** This sub menu consist of four buttons: “SELECT” for selecting the title of E-book/E-Comic, “BACK” for return to main menu, “PREVIOUS” to go to previous page, and “NEXT” to go to next page of opened content. Before user could read the content, user have to select the title of the E-book/E-Comic by navigating “PREVIOUS” and “NEXT” button. After the desired title is shown, user opens it by selecting the “SELECT” button and the content of selected file will be opened and shown on display.

The sub menu for internet browsing is shown in Figure 5. This sub menu will allow user surfing around the world through web site. User could use our interface for browsing internet by utilizing his eye only. First, user input the URL address via moving keyboard navigated using four buttons: “UP” is for moving the layout go to upward, “DOWN” is for moving the layout go to downward, “LEFT” is for moving the layout go to leftward, and “RIGHT” is for moving the layout go to rightward. After the URL address is input by user, the web page will be shown on bottom part of our interface.

The last sub menu is TV Information Extraction is shown in **Error! Reference source not found.** It will allow user extract information from Digital TV (Usually used to extract information such as schedule, price of advertising item, sub title, etc). To extract the information, user could navigate our interface using four buttons: “LEFT” and “RIGHT” are for changing the type of information, “BACK” is for returning to main menu, and “EXTRACT” is for executing the TV Information follows the type of information.



Figure 4 Sub menu of E-book/E-comic/E-Learning content reader



Figure 5 Sub menu of Internet browsing

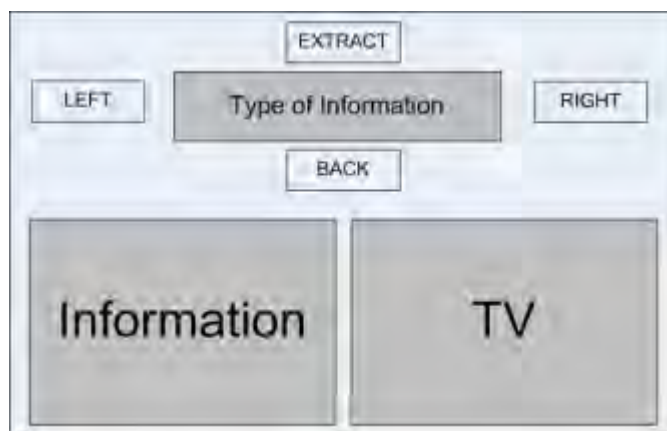


Figure 6 Sub menu of TV information extraction

B. Implementation

We implement our system by utilizing Infrared: IR Camera, NetCowBoy DC NCR-131 mounted on user glasses to acquire user image. We modified the position of 7 IR LED for adjusting illumination and obtaining stable image even illumination of environment changes as shown in Figure 7.

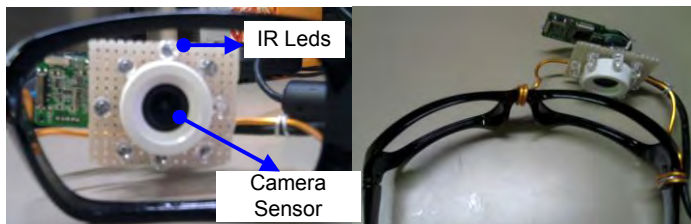


Figure 7 Modified camera sensor

Our software is developed under C++ language of Visual Studio 2005 and OpenCv, image processing library, which can be downloaded as free at their website. The advantages of this

camera mounted on glasses have been explored in ref [4]. It was success to minimize problems such as vibration, illumination changes, head movements, etc.

In this system, we search pupil location on an eye image by using our method that has been published in the reference [16]. We estimate the sight by converting the obtained position of pupil to sight angle. After the sight angle is estimated, we control the mouse cursor by using this sight.

The use of typical web camera has merit in low cost and easy to make, unfortunately it has demerit in noise, flicker, low resolution, etc. These demerits influence the stability of our system. Also, the various type of user's eyelash, deformable phenomenon of eye shape due to eye movement, existence of light source, etc often influence our sight result become unstable.

To solve this stability problem, there are many approach such as improving hardware stability, filtering, etc. In this system, we solve this problem by developing interface allowing user type characters; navigate an E-book/E-Comic reader, etc. We maintain the typing accuracy by developing the interfaces that have been explained previously.

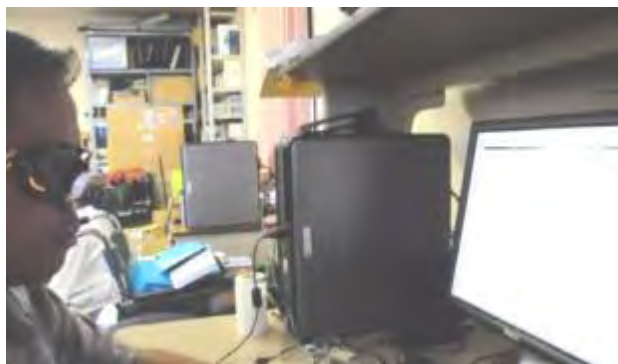


Figure 8 Use of the proposed system

To use our system, user just wears the glasses sensor in front of display as shown in Figure 8. Before user uses our system, the calibration step should be passed first (because we only use single camera). The calibration step will synchronize eye trajectory on image (which is acquired by camera) with the sight trajectory on the display. The eye trajectory in image has different pattern compared with eye trajectory in display. Due to the difference of camera placement, the sight estimated result may have nonlinear output with display as shown in Figure 9.

If the camera placement is not in center of pupil exactly, it means the plane of camera is not in a line between center of pupil and center of display, we have to transform the trajectory output of camera to trajectory of display. The calibration points that resulted from calibration step influenced by the different camera placement are shown in Figure 10.

From Figure 10 we can see that between target points on

display has different pattern to calibration points on camera image. We can see that the calibration points is little bit nonlinear and symmetry. To solve this non linearity problem, we use perspective transformations to transform the nonlinear eye estimated location to trajectory of display as shown in Figure 11.

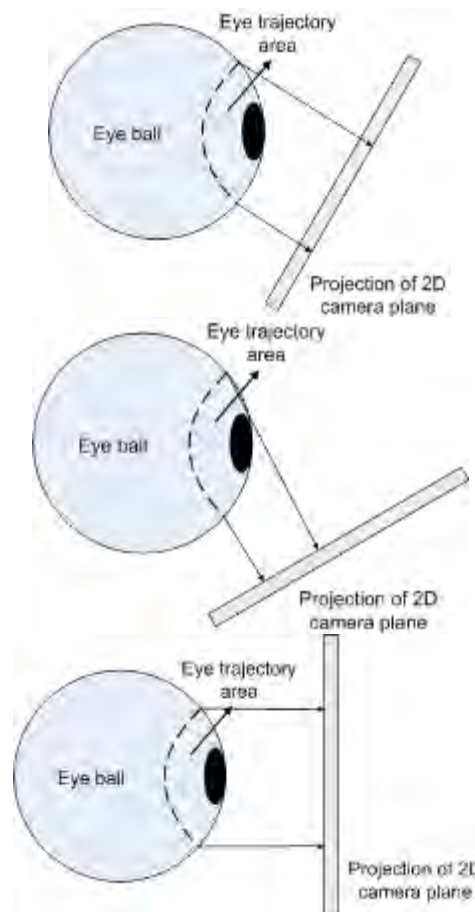


Figure 9 Phenomena of different camera placement to the relation between eye trajectories to display-trajectory area

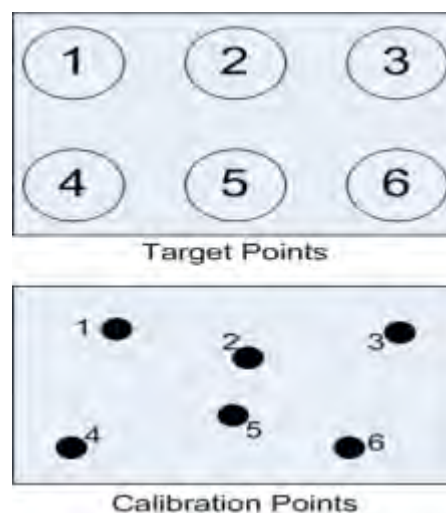


Figure 10 Effect of different camera placements

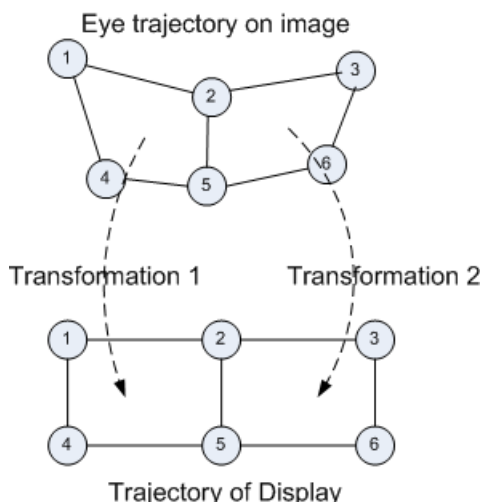


Figure 11 Transformation from eye trajectory to display trajectory

III. EXPERIMENTAL RESULTS

To measure the effectiveness of our proposed method, we tested its performance by conducting accuracy experiment for eye detection and the sight.

A. Eye Detection Accuracy

The perfect performance of eye detection method is mandatory in every eye-based HCI system. It determines of accuracy in early step. The next process only will gain the result. This experiment involved six users with different nationality (The difference of nationality identical with various eye shapes, various skin colors of eye, etc). The TABLE 1 shows the accuracy performance of our method compared with two other methods: Adaptive Threshold⁹ and Template Matching¹⁰. The result shows that our method is superior with accuracy of 96.73% and could maintain accuracy against different user by variance is 16.27%.

Also, we test our method to light changing. We give adjustable lighting to system and measure the effect. The result shows that our method could maintain the accuracy from 0 Lx until about 1500 Lx. It means that our method does work without any light because we have IR LED to adjust the illumination. The proposed method failed if the light of environment is more than 2000 Lx (direct sun light).

TABLE 1. EYE DETECTION ACCURACY

User Types	Nationality	Adaptive Threshold (%)	Template Matching (%)	Proposed Method (%)
1	Indonesian	99.85	63.04	99.99
2	Indonesian	80.24	76.95	96.41
3	Sri Lankan	87.8	52.17	96.01
4	Indonesian	96.26	74.49	99.77
5	Japanese	83.49	89.1	89.25
6	Vietnamese	98.77	64.74	98.95
Average		91.07	70.08	96.73
Variance		69.75	165.38	16.27

⁹ <http://homepages.inf.ed.ac.uk/rbf/HIPR2/adpthrsh.htm>

¹⁰ http://en.wikipedia.org/wiki/Template_matching

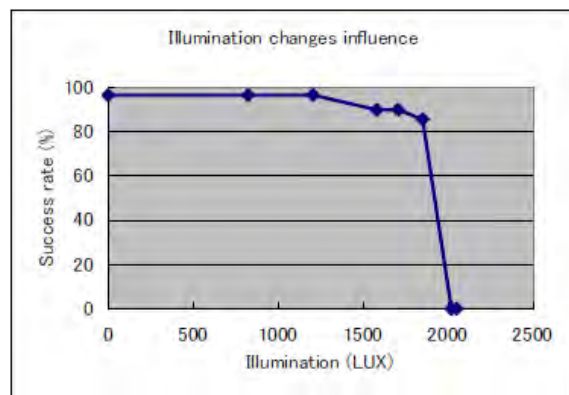


Figure 12 Influence of light changing to eye detection accuracy

B. Stability of Sight Estimated Point

The sight stability is measured to know the radius of sight error. This radius determines the maximum number of key that still could be used. The bigger radius of sight error causes the maximum number of key become decreases. Otherwise, the minimum radius of sight error could elevate the maximum number of key. In this experiment, user was looking at the target point and it moved serially on six locations. The Figure 13 is shown the sight stability. It shows that on key 5, the radius of sight error is high compared with other keys (it was caused by light source disturbed the eye detection method).

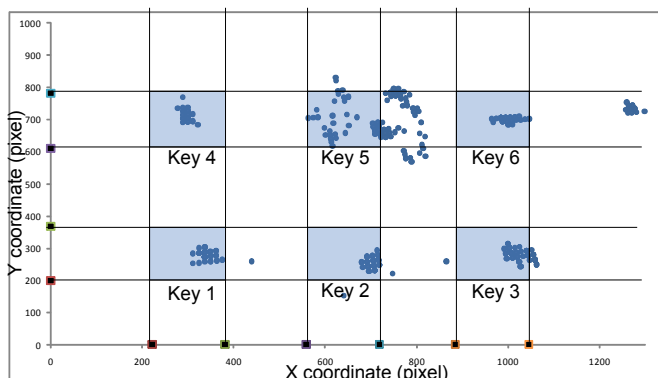


Figure 13 Sight Stability

C. Effect of High Number of Key to Accuracy

This experiment measured the effect of high number of key to accuracy. We conduct the experiment by modifying the number of key and measuring the accuracy. We start our experiment by using the number of key is 2; continue with 3, 5, 7, 9, and 15. The effect of number key to accuracy is shown in Figure 14. It shows that the raise of number of key will increase the error (decrease the accuracy). Also, we made a simulation that could figure the relation of accuracy to the distance among keys. This relation could be draw to a model as shown in Figure 15, with the assumptions are sight instability follow circle distribution (non parameter) with the radius is R, the distance among key is AB, and the error is represented by the slices among the circles.

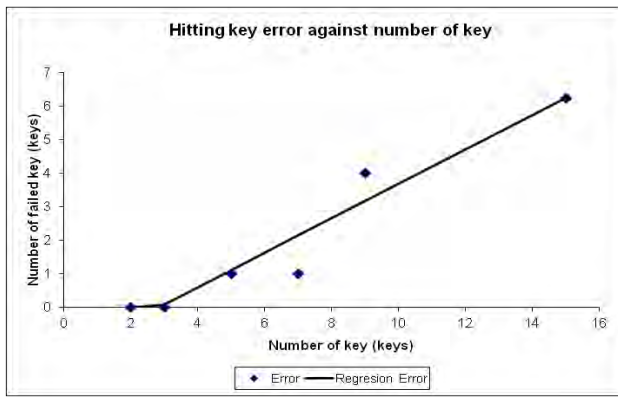


Figure 14 Effect of number of keys to accuracy

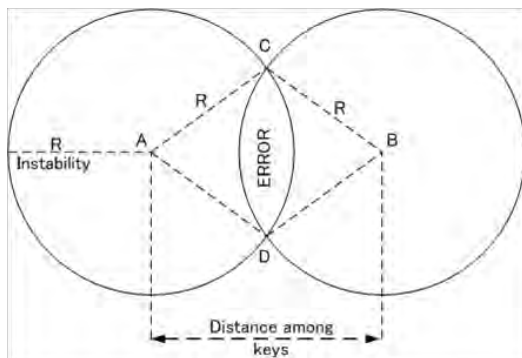


Figure 15 Model of relation accuracy to distance among keys

The result of simulation is shown in Figure 16. It shows that the approach of distance among keys causes the accuracy decreases. Otherwise, the widening of distance among keys causes the accuracy become maximal.

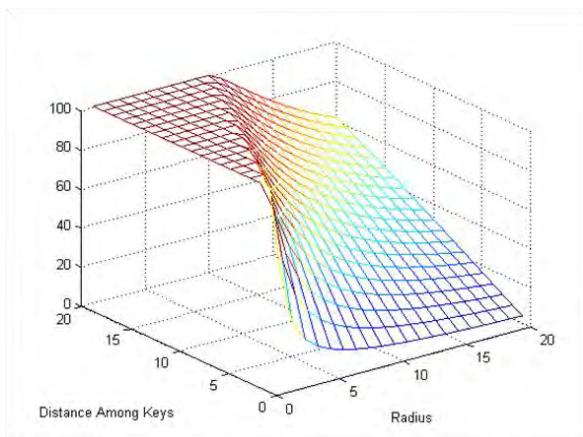


Figure 16 Simulation result of relation accuracy to distance among keys

D. Key-in Accuracy and Process Speed

In the last experiment, we conducted the experiment for measuring typing accuracy and speed. In this system, we use several model interfaces including standard navigator interface with five key, single line of moving keyboard, and multi-line of moving keyboard. In this experiment, we measured the typing accuracy when user was using multi line of moving keyboard and also recorded the typing speed.

In this experiment, we involved six users including two beginner users and four expert users. The beginner user means they ever used our system less than ten times. Otherwise, the expert user ever used our system more than ten times. How much time they ever used our system determines their expertise level. We compared our result to Fixed Keyboard Model as shown in TABLE 2.

TABLE 2 THE ACCURACY OF MULTI LINE OF MOVING KEYBOARD

User	Expertise	Moving Keyboard (%)	Fixed Keyboard (%)
1	Expert	100.00	92.86
2	Expert	100.00	76.19
3	Beginner	82.14	71.43
4	Expert	100.00	85.71
5	Beginner	71.43	78.57
6	Expert	100.00	66.67
Average		92.26	78.57

The experiment result shows that our system has better accuracy compared with fixed keyboard model. It was caused by our key was bigger than the fixed keyboard. Also, we only used five keys to navigate the moving keyboard while the used fixed keyboard in this experiment has thirty keys. According to our simulation result in Figure 16 that the higher number of key will has worse accuracy. Because our method used lower number of key, it causes our accuracy better than the fixed keyboard model.

Beside the measurement of typing accuracy, we measured the typing speed. By using same methodology of typing accuracy experiment, we recorded the typing speed. The experiment result is shown in TABLE 3. It shows that our method is faster than the fixed keyboard model because the use of smaller key (in fixed keyboard model) made user become difficult to input a character and it made the typing speed become slower. Otherwise, our method used bigger key and it made user still possible input a character easily. The result shows that our method is faster with the typing speed of 134.69 seconds while the fixed keyboard has slower typing speed of 210.28 seconds.

TABLE 3 TYPING SPEED

User	Expertise	Moving Keyboard (s)	Fixed Keyboard (s)
1	Expert	117.50	154.00
2	Expert	138.67	195.33
3	Beginner	180.50	275.00
4	Expert	101.00	197.33
5	Beginner	161.50	213.00
6	Expert	109.00	227.00
Average		134.69	210.28

E. All the Functionalities

All the functionalities, phoning, reading E-Book, E-Learning, E-Comic contents, Internet browsing, watching TV and the required information for purchasing products extraction from TV commercial are confirmed.

IV. CONCLUSION

It is concluded that the proposed eye-based HCI system works well for selecting and determining five keys for navigation of functionalities. Expertise is required for perfect key-in. In other words, 100% of key-in success rate can be achieved through exercises of using eye-based HCI system. Comparative study between the conventional fixed keyboard and the proposed moving keyboard shows that key-in speed of the proposed system is much faster than that of the conventional system by around 35%. All the functionalities, phoning, internet browsing, reading E-book/E-Comic/E-learning contents, and TV information extraction are confirmed. These functions are available in a real time basis.

ACKNOWLEDGMENT

Authors would like to thank to the graduate students who contributes to the performance evaluation experiments of the proposed system.

REFERENCES

- [1] FOUNDATION, Christopher and Dana Reeve. Disponivel em: <<http://www.christopherreeve.org>>. Acesso em: 7 October 2011.
- [2] EYEWRIter: low-cost, open-source eye-based drawing system. Disponivel em: <<http://www.crunchgear.com/2009/08/25/%20eyewriter-low-cost-%20open-source-eye-%20based-drawing-system/>>. Acesso em: 7 October 2011.
- [3] LEFF, R. B.; LEFF, A. N. 4.954.083, 1990.
- [4] Arai, K.; R. Mardiyanto, A Prototype of ElectricWheelchair Controlled by Eye-Only for Paralyzed User. Journal of Robotic and Mechatronic, 23, 1, 66-75, 2011.
- [5] Djoko P., R.Mardiyanto, and K.Arai, Electric wheel chair control with gaze detection and eye blinking. Artificial Life and Robotics, AROB Journal. 14, 694,397-400, 2009.
- [6] Arai K; K. Yajima, Robot Arm Utilized Having Meal Support System Based on Computer Input by Human Eyes Only. International Journal of Human Computer Interaction (IJHCI), 2, 1, 120-128, 2011.
- [7] EYE tracking and eye Control. Disponivel em: <<http://www.tobii.com/>>. Acesso em: 8 out. 2011.
- [8] OFFICIAL website of Professor Stephen W. Disponivel em: <<http://www.hawking.org.uk>>.
- [9] John J. M. et al. EyeKeys: A Real-Time Vision Interface Based on Gaze Detection from a Low-Grade Video Camera. 2004 Conference on Computer Vision and Pattern Recognition Workshop. 159, 2004.
- [10] Changzheng L., K. Chung-Kyue, P.Jong-Seung, The Indirect Keyboard Control System by Using the Gaze Tracing Based on Haar Classifier in OpenCV. Proceedings of the 2009 International Forum on Information Technology and Applications. 362-366, 2009.
- [11] Zhu H, L. Qianwei, Vision-Based Interface: Using Face and Eye Blinking Tracking with Camera. Proceedings of the 2008 Second International Symposium on Intelligent Information Technology Application. 306-310, 2008.
- [12] Parks K.S., L. K. T. Eye-controlled human/computer interface using the line-of-sight and the intentional blink. Computers and Industrial Engineering. 463-473, 1993.
- [13] Barea, R. et al. System for Assisted Mobility using Eye Movements based on Electrooculography. IEEE Transaction on Neural System and Rehabilitation Engineering, 10, 209-218, 2002.
- [14] Arai K. and T. Herman, Automatic e-comic content adaptation, International Journal of Ubiquitous Computing, 1,1,1-11,2010.
- [15] Arai K., and T. Herman, Method for extracting product information from TV commercial, International Journal of Advanced Computer Science and Applications, Special Issue on Artificial Intelligence, 2, 8, 125-131, 2011.
- [16] Arai, K.; R.Mardiyanto, Improvement of gaze estimation robustness using pupil knowledge. Proceedings of the International Conference on Computational Science and Its Applications (ICCSA2010). 336-350, 2010.

AUTHORS PROFILE

Kohei Arai received a PhD from Nihon University in 1982. He was subsequently appointed to the University of Tokyo, CCRS, and the Japan Aerospace Exploration Agency. He was appointed professor at Saga University in 1990. He is also an adjunct professor at the University of Arizona since 1998 and is Vice Chairman of ICSU/COSPAR Commission A since 2008

Very Low Power Viterbi Decoder Employing Minimum Transition and Exchangeless Algorithms for Multimedia Mobile Communication

Prof. S. L. Haridas

HOD Electronics & Telecommunication Engineering
B. D. College of Engineering, Sevagram
Wardha, M.S., India

Dr. N. K. Choudhari

Principal
Smt. Bhagvati Chaturvedi College of Engineering
Nagpur, M.S., India

Abstract— A very low power consumption viterbi decoder has been developed by low supply voltage and 0.15 μm CMOS process technology. Significant power reduction can be achieved by modifying the design and implementation of viterbi decoder using conventional techniques traceback and Register Exchange to Hybrid Register Exchange Method (HREM), Minimum Transition Register Exchange Method (MTREM), Minimum Transition Hybrid Register Exchange Method (MTHREM), Register exchangeless Method and Hybrid Register exchangeless Method. By employing the above said schemes such as, HREM, MTREM, MTHREM, Register exchangeless Method and Hybrid Register exchangeless Method; the viterbi decoder achieves a drastic reduction in power consumption below 100 μW at a supply voltage of 1.62 V when the data rate of 5 Mb/s and the bit error rate is less than 10^{-3} . This excellent performance has been paved the way to employing the strong forward error correction and low power consumption portable terminals for personnel communication, mobile multimedia communication and digital audio broadcasting. Implementation insight and general conclusions can particularly benefit from this approach are given.

Keywords- Hybrid register exchange method; minimum transition register exchange method; minimum transition hybrid register exchange method; register exchangeless method; hybrid register exchangeless method.

I. INTRODUCTION

The convolutional encoding and viterbi decoding schemes [1] is used by many mobile communication, satellite communication or broadcasting systems [2, 3]. Because the scheme shows powerful forward error correction performance and the great progress in CMOS technology makes it possible to realize the high speed, low power encoders/decoders [4]. A viterbi decoder [5, 6] is an important target for power reduction in many low power communication devices such as cellular phones, where it consumes almost one third power [7]. Higher memory order codes can achieve superior performance without requiring additional channel bandwidth. However, to counteract the exponential dependence of viterbi decoder complexity on memory order in low power designs, good power reduction methods are needed. Continuous efforts by defining various power reduction algorithms or approach such as Hybrid Register Exchange Method (HREM), Minimum

Transition Register Exchange Method (MTREM), Minimum Transition Hybrid Register Exchange Method (MTHREM), Register exchangeless Method, and Hybrid Register exchangeless Methods are given by the author in [8, 9, 10, and 11].

As in the case of memory designs today, the significant power reduction potential lies in the dynamically varying a viterbi decoder implementation according to real time changes in system characteristics. The goal of the approach proposed in this paper is to reduce power consumption while decoding convolutional codes in a system where acceptable bit error rate varies in real time. There are following main contributions of this paper.

- 1) A system dependent, low power approach for decoding convolutional codes namely Hybrid Register Exchange Method (HREM), Minimum Transition Register Exchange Method (MTREM), Minimum Transition Hybrid Register Exchange Method (MTHREM), Register exchangeless Method and Hybrid Register exchangeless Methods are demonstrated.
- 2) Variation in the potential of these approaches as system characteristic maximum acceptable bit error rate (MABER) varies is studied.
- 3) Comparisons of power reduction potential of above mentioned approaches are demonstrated.

II. BACKGROUND

For the purpose of this work, a viterbi decoder implementation that employs register exchange can be described in terms of memory 'm', registers used for storing the survivor memory and free distance of convolutional code. A Minimum transition and Exchangeless algorithm decoders can be described with one additional characteristic, threshold. Viterbi decoding with Register Exchange method is performed by using the bits received to generate the path that represents likely transitions made by the convolutional encoder state machine over time and transfer the likely data bits to every state registers at every time. The convolutional encoder memory order determines the height or number of states per stage of the trellis, which represents the number of paths stored at any time by the viterbi decoder.

The proposed approaches, referred to as minimum transition and exchangeless decoding algorithms reduces power consumption by adopting threshold employing register exchange method to real time system changes; such as the maximum acceptable bit error rate (MABER). These approaches greatly impart power consumption because threshold controls the average number of paths stored per trellis stage and number of registers required to store these paths. The number of operations performed by the decoder, especially the memory access and ultimately the switching activity is highly dependent on the average number of path stored and the number of states (register used for storing path). Maintaining the Integrity of the Specifications.

III. SYTEM ARCHITECTURE

A Viterbi decoder and a convolutional encoder operate by finding the most likely decoding sequences for an input code symbol stream. A convolutional encoder is selected for error correction with digital mobile communication. We consider a convolutional code (3, 1, and 4) which has a code rate 'r' of 1/3 and constraint length 'K' of 5 and number of memory element 'm' is 4. The shift register taps terminate at modulo-2 adders forming generator function ($g_2 = 25, g_1 = 33, g_0 = 37$). Input bit enters the shift register by one bit at a time. The outputs of the generator functions produce an output of three symbols for each input bit, which corresponds to a code rate of 1/3. The implementation of the viterbi algorithm referred to as the viterbi decoder. The major blocks of a viterbi decoder are shown in Fig. 1. The role of each block is described briefly below.

- Butterfly unit: In butterfly we combine Branch Metric computation and Add Compare Select (ACS) block, which is first calculate the branch metric, then it get added to state metric to get path metric. Then compare the two path metrics and select the lower path metric path which is a survivor for that state.
- Survivor path storage: it stores the survivor path metric for further computation and decision regarding whether it is lesser than threshold or not.
- Output decoding block: it produces decoded bits.
- Control unit: it provides the synchronization and initiates the control signals to all blocks.

IV. DECODING METHODS

There are two standard methods used in viterbi decoder to decode the data bits; namely register exchange and traceback methods. We employ the register exchange method and modified it for power reduction.

A. Register Exchange Method

In Register Exchange Method, a register is assigned to each state which contains information bits for the survivor path throughout the trellis. Total number of states and the corresponding register requires are 2^m . The register keeps the partially decoded output sequence along the path. The register exchange method eliminates the need to traceback since the register of final state which is same as initial state S_0 contains the decoded output. This approach results in complex hardware

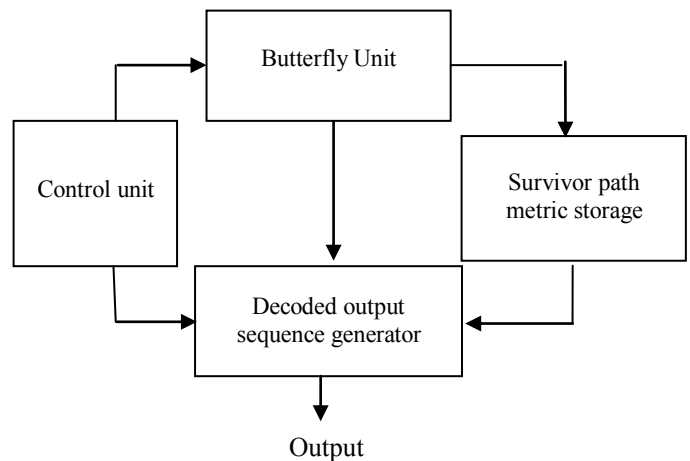


Figure 1. Block diagram of viterbi decoder

and high switching activity due to the need to copy the content of all the registers from state to state at every time. So the power consumption of viterbi decoder using REM is very high.

B. Hybrid Register Exchange Method

In this method, register exchange and traceback methods are combined and therefore the names Hybrid register exchange method [8]; which reduces the switching activity and power. Here, we are using a property of trellis is that, for m cycles the data bits will be the corresponding present state bits irrespective of the initial state from where the data gets transferred? To find the initial state we have to traceback through an 'm' cycles by observing the survivor memory, and then transfer the partial decoded data from an initial state to the next state which is m cycle later (not at subsequent cycle as in REM) and the present decoded data will be the present state itself. The memory operation is not in every cycle, and it gets reduced by a factor of m. Also, the shifting of data from one register to another is reduced that is the switching activity and ultimately the power consumption will get reduce.

C. Minimum Transition Algorithms (MTREM & MTHREM)

The main drawback of register exchange method is its frequent switching activity. One of the promising solutions to reduce the switching activity is to avoid unwanted data transfer from one register to other at every time interval. The path metric at every time interval describes, how much bits they are differed from the transmitted sequence. In these minimum transition algorithms we use the property of free distance of the encoder. Free distance tells about the errors which can be corrected by the viterbi decoder. We setup these errors as a threshold since above which the error cannot be corrected, so there is no need for storing the survivors which has a path metric above this threshold.

At the i^{th} decoding time stage, the state who has path metric greater than the threshold are eliminated in data exchange process, where threshold is fixed value depending on free distance of encoder. Here, we require only half of the registers than simple register exchange and hybrid register exchange method i.e. $2^{m/2}$. This is because there are maximum $2^{m/2}$ states which have path metric lesser or equal to the threshold. We can avoid these states or path from data transfer, so that the

data bits transitions could not take place and avoid undesirable switching. Also, these state registers do not contain any further data transition from these states at that time and hence further switching reduces. One of such approaches is Minimum Transition Register Exchange Method (MTREM).

In Hybrid register exchange method, data transfer is taking place at m^{th} instant of time therefore switching activity is less. Here, also fix a threshold and avoid an unwanted state register and data transfer operations. At m^{th} instant of time the current data bits are the current state bits only. Here the all operations i.e. data transfer and traceback operations are placed at m^{th} instant, switching activity and power dissipation further reduce.

D. Exchangeless RE and HRE Algorithms

In register exchange method, hybrid register exchange method and minimum transition methods; each row of memory is used to trace the decoded bits, if an initial state is assumed. The bottom row register in all above methods consist an decoded data, if an initial state is S_0 . Further if the initial and final states are same, then the extra rows of memory are not needed. If it is assumed that initial and final state is S_0 , then only one row corresponding to S_0 is required. If we observe carefully, the path metric at every instant of time, there is only one state who has path metric less than or equal to one. Therefore if we put a threshold one then at every instant there will be only one state where data to be transferred. For the even state data is '0' and for odd state it becomes '1'. If we continue in this fashion, after the reset sequence the data at the final instant is decoded data. We can say it is a register exchangeless method. Similarly for hybrid register exchangeless method, data will be decoded after every 'm' clock cycle, where the data is only the address of that state.

The exchangeless viterbi decoder is a low power design for the viterbi decoder with the restriction of resetting the encoder register after each 'L' (survivor path length) encoded data bits and it can be used for a bit error rate of acceptable limit 10^{-5} to 0.03 for wireless applications with the assumption that there will be no consecutive errors.

V. EXPERIMENTAL RESULT

In this section, the power reduction potential of the above mentioned algorithm decoder is accessed through experimental results. For a system, with variable maximum acceptable bit error rate which is applicable when minimum error correction performance needs vary over time because of variation in the type of information being received; is experimented and corresponding power reduction is mentioned. The reason for power consumption can be significantly reduced for algorithms by reducing threshold is mainly because it affects the number of times path memory is accessed. Actually the average number of times nearly all calculations are performed per trellis stage is proportional to the average number of surviving paths per stage which is controlled by the threshold. Also the average number of memory read and write operations performed per stage is proportional to the number of state registers that are accessed during data exchange between registers.

The decoder of rate 1/3 with constraint length K of 5 has been implemented using the 0.15 μm CMOS technology which operates at a supply voltage of 1.62 V. The power dissipation

of the developed decoder with various above said algorithms has been measured and plotted in fig. 2. Power dissipation of the viterbi decoder using register exchange, hybrid register exchange, minimum transition register exchange, minimum transition hybrid register exchange, register exchangeless and hybrid register exchangeless algorithms are estimated as 212.26 μW , 156.35 μW , 87.41 μW , 76.75 μW , 54.95 μW and 52.13 μW respectively at a bit error rate of 10^{-6} .

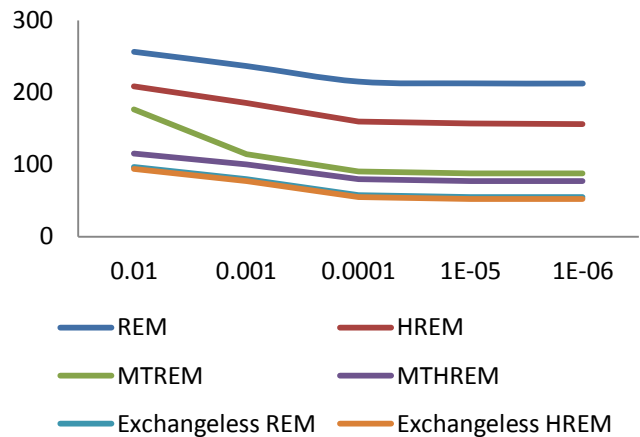


Figure 2. Power Consumption (μW) vs. BER

VI. CONCLUSION

A very low power consumption viterbi decoder is implemented with 0.15 μm technology. By applying the exchangeless hybrid register exchange method, the decoder with rate 1/3 and constraint length of 5, achieved drastic power reduction. The power consumption is reduced by almost 75% at bit error rate of 10^{-6} when the data rate is 5 Mb/s. the data rate is large enough to be implemented into the portable terminal of the audio broadcasting satellite service, mobile multimedia communication and so on.

REFERENCES

- [1] Viterbi A. J., "convolutional codes and their performance in communication systems", IEEE Trans., COM-19, pp. 761-772, 1967.
- [2] Kato S., Morikura M., Kubota S., Kazama H. and Enomoto E., "A TDMA satellite communication system for ISDN services", Globecom'91, pp. 1533-1540, 1991.
- [3] Kameda K., "audio broadcasting satellite service via communication satellites", NHK R&D, pp. 9-17, 1992.
- [4] Kawazoe K. Honda S., Kubota S. and Kato S., "ultra high speed and universal coding rate viterbi decoder VLSIC", Proc. ICC'93, pp. 1434-1438, 1993.
- [5] Viterbi A. J., "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm", IEEE Trans. On Information theory, vol. IT-13, pp. 260-269, 1967.
- [6] Forney G. D., Jr., "The viterbi algorithm", Proc. IEEE, vol. 61, no. 3, pp 218-278, 1973.
- [7] Kang I. and Wilson A. Jr., "low power viterbi decoder for CDMA mobile terminals", IEEE J. of solid-State Circ., vol. 33, no. 3, pp. 473-482, 1998.
- [8] Haridas S. L. and Dr. Choudhari N. K., "Design of Viterbi decoder with modified Traceback and Hybrid Register Exchange processing" Proc. ACM Int. Conf. ICAC3'09, pp. 223-226, 2009.
- [9] Haridas S. L. and Dr. Choudhari N. K., "A low power viterbi decoder design using hybrid register exchange processing for wireless

- applications”, *Int. J. of Electronics & Telecommunication and Instrumentation Engineering (IJETIE)*, vol. 03, no. 1, pp. 15-20, 2010.
- [10] Haridas S. L. and Dr. Choudhari N. K., ”A low power viterbi decoder design using minimum transition hybrid register exchange processing for wireless applications”, *Int. J. VLSI design and communication system(VLSICS)*, vol.1, No. 4, pp 14 to 22, 2010.
- [11] Haridas S. L. and Dr. Choudhari N. K., “A low power viterbi decoder design using Register exchangeless Approach for wireless applications”, paper submitted to IETE journal of research.

AUTHORS PROFILE

Prof. S. L. Haridas received Bachelor of Engineering and Master of Engineering degrees from Nagpur University, M.S., India in 1988 and 1995

respectively. Currently he is pursuing the Ph. D. degree at R. T. M. Nagpur University. He is a Professor and Head of Electronics and Telecommunication Engineering department at B. D. College of Engineering, Sevagram, M. S., India.

Dr. N. K. Choudhari received Bachelor of Engineering and Master of Engineering degrees from Nagpur University, M.S., India in 1987 and 1993 respectively. He received Ph. D. degree from J. M. University, New Delhi, India. Presently he is working as a Principal at Smt. Bhagvati Chaturvedi College of Engineering, Nagpur, M. S., India. He published almost 30 research papers in various conferences and reputed journals. He held different positions on various committees and bodies at university level. He guided nearly 10 research scholars in electronics engineering.

Outlier-Tolerant Kalman Filter of State Vectors in Linear Stochastic System

HU Shaolin

State Key Laboratory of Astronautics
Xi'an, 710043, P.O.Box 505-16, China;

Karl Meinke

CSC School, Royal Institute of Technology
SE-100 44, Stockholm, Sweden

Huajiang Ouyang

School of Engineering, The University of Liverpool
Liverpool, L69 3GH, UK;

SUN Guoji

System Engineering Institute, Xi'an Jiaotong University
Xi'an, 710043, China

Abstract—The Kalman filter is widely used in many different fields. Many practical applications and theoretical results show that the Kalman filter is very sensitive to outliers in a measurement process. In this paper some reasons why the Kalman Filter is sensitive to outliers are analyzed and a series of outlier-tolerant algorithms are designed to be used as substitutes of the Kalman Filter. These outlier-tolerant filters are highly capable of preventing adverse effects from outliers similar with the Kalman Filter in complexity degree and very outlier-tolerant in the case there are some outliers arisen in sampling data set of linear stochastic systems. Simulation results show that these modified algorithms are safe and applicable.

Keywords- Kalman filter; Outlier-tolerant; Outlier; Linear stochastic system.

I. INTRODUCTION

The Kalman filter is not only a widely used tool to estimate or to reconstruct states of a dynamic system in modern control but also famous powerful tool to extract useful information from noisy signals in signals processing. The Kalman filtering algorithms have many advantages: it is optimal in linear estimator set and suitable for online processing because of its recursive relationship; it can be used in stationary system and non-stationary system; and it can be used in multi-dimensional processes, etc. Since Kalman and Bucy put forward this linear optimal iterative filtering algorithm with the development of state-space theory in the early 1960s', its applications have become more and more wide-spread in many different engineering fields, such as process control, stochastic control and navigation, etc. It is also very useful in fault diagnosis of dynamic systems.

Although the Kalman filter possesses many advantages stated above, recent research has revealed that the Kalman filtering algorithm is not robust (see [1,3]) against perturbation in a model or observed/measured data. Practical experience in using Kalman filter to process signals also indicates that outliers in the measured data would degrade the performance of Kalman filter. How to improve the Kalman filtering algorithm is an open question in the cases when there are outliers in measurement data sequence because outliers are unavoidable

and could lead to a considerable deviation of the estimated target from the true system status when using Kalman-filter based algorithms. Durovic, et al. (1999) discussed robust estimation with unknown noise statistics; Nihal et al. (1991) and Chan et al. (2005) built a new robust Kalman filter algorithm with outliers respectively; Ting et al. (2007) reviewed the Kalman filter and suggested a kind of robust Kalman filtering with Bayesian weights so as to overcome negative effects from outliers.

This paper analyzes systematically the adverse effect of outliers on Kalman filter and establishes a series of outlier-tolerant filtering algorithms. Numerical results of simulated examples show the validity of the outlier-tolerant algorithms.

II. EFFECTS OF OUTLIERS ON KALMAN FILTER

The Kalman filter is used to estimate the state vector of the following linear stochastic dynamic-measurement system

$$\begin{cases} X_{k+1} = A_{k+1}X_k + \varepsilon_k \\ Y_{k+1} = H_{k+1}X_{k+1} + \eta_{k+1} \end{cases} \quad (X_k \in R^n, Y_k \in R^m) \quad (1)$$

If both the multi-dimensional dynamic noise process $\{\varepsilon_k\}$ and dynamic measurement error series $\{\eta_k\}$ possess the following properties:

- (a) $E\{\varepsilon_k\} = E\eta_k = 0$
- (b) $\text{cov}(\varepsilon_k, \eta_k) = 0$
- (c) $R_{\eta(k)} = \text{cov}(\eta_k, \eta_k)$, $R_{\varepsilon(k)} = \text{cov}(\varepsilon_k, \varepsilon_k)$

then the optimal estimator of the state vector X_k in model (1) can be expressed by the following recursive relationships:

$$\begin{cases} \hat{X}_{(k+1|k+1)} = A_{k+1}\hat{X}_{(k|k)} + K_{k+1}\hat{E}_{(k+1|k)} \\ \hat{X}_{(k+1|k)} = A_{k+1}\hat{X}_{(k|k)} \\ K_{k+1} = \sum_{k+1} H_{k+1}^T \{H_{k+1} \sum_{k+1} H_{k+1}^T + R_{\eta(k+1)}\}^{-1} \\ \sum_{k+1} = A_{k+1}(I - K_k H_k) \sum_k A_{k+1}^T + R_{\varepsilon(k)} \end{cases} \quad (2)$$

These recursive relationships can intuitively express that the best estimator $\hat{X}_{(k+1|k+1)}$ of the state X_{k+1} at time t_{k+1} is composed of two parts: the best estimator of the state X_k at time t_k and the sampling innovation as follows

$$\hat{E}_{(k+1|k)} = Y_{k+1} - H_{k+1}A_{k+1}\hat{X}_{(k|k)} \quad (3)$$

Obviously, if the additional samples Y_{k+1} are “normal” values, the sampling innovation $\hat{E}_{(k+1|k)}$ will make some correct modification to predictor $\hat{X}_{(k+1|k)}$ with ratio K_{k+1} to get the next optimal estimator $\hat{X}_{(k+1|k+1)}$. On the other hand, if the additional samples Y_{k+1} are outliers, the resultant innovation $\hat{E}_{(k+1|k)}$ will be abnormal and the abnormal information will result in an erroneous modification to prediction $\hat{X}_{(k+1|k)}$ by the same ratio K_{k+1} , which lead to the filtering estimators deviating from normal states.

In section V of this paper, an example is given to substantiate the negative influence of outliers on the Kalman filtering estimators of state vectors. Figure 1 plots the simulation results, which indicates that the Kalman filtering of state vectors is far from normal states when there are outliers in sampling data series.

For a practical measurement system or device, outliers are inevitable in sampling data because of faulty operations or recording errors. In some cases, there may be complicated abnormal measurement data existing in sampling processes, such as step-type jumps or patchy outliers, etc. For example, when the flight of a spacecraft was tracked by an impulse radar, 1%~2% (occasionally as high as 5%) of measurement data display serious deviations from the trend formed by most other samples. So it is very important to modify the Kalman filtering algorithms or reduce the negative effects of outliers on Kalman filtering estimators of state vectors.

III. TACTICS TO DEAL WITH OUTLIERS IN SAMPLE

Considering that it is very difficult to diagnose outliers in a large quantity of data and that the Kalman filtering is still widely used, a best effort is made in this paper to improve fault-tolerance of the linear optimal recursive filtering algorithm. The improved filters should possess the following properties:

- 1) The algorithm should be recursive and easy to use;
- 2) When there are a few abnormal samples in the measured data series, the filtering algorithm must have a strong ability to overcome negative effects from abnormal data, or must restrict the negative effects within prescribed bounds;
- 3) When there are no abnormal samples, the algorithm is capable of making full use of useful information to achieve high filtering accuracy.

Considering these three restrictive conditions stated above, a set of algorithms which are similar to the Kalman filter are established as follows

$$\begin{cases} \tilde{X}_{(k+1|k+1)} = A_{k+1}\tilde{X}_{(k|k)} + K_{k+1}\Phi_{k+1}(r_{k+1})\hat{E}_{(k+1|k)} \\ r_{k+1} = \hat{E}_{(k+1|k)}^T G_{k+1}^{-1} \hat{E}_{(k+1|k)} \end{cases} \quad (3)$$

where the function series $\{\Phi_{k+1}(\cdot)\}$ are segment-wise smooth and bounded, K_{k+1} is the gain and G_{k+1} is a weighting matrix.

When $\Phi_{k+1} = 1$ and $G_{k+1} = I$ (identity matrix) are selected, $\tilde{X}_{(k+1|k+1)}$ in equation (3) is reduced to the conventional Kalman filter. It is found that the main reason for outlier-tolerance of conventional Kalman filter is that the function sequence $\{\Phi_{k+1} = 1\}$ potentially treat all of the innovations (normal and abnormal) equally. This is the root cause why Kalman filter is unable to deal with outliers. In order to endow a filter anti-outliers capability, a sensible tactic is to select a suitable Φ_{k+1} which must decrease or diminish to zero when r_{k+1} increases.

IV. OPTIMAL SELECTION OF Φ_k

In order to reduce the negative effects of outliers in state filtering, some analyses of formulae (3) must be done as follows:

$$\begin{aligned} \Delta\tilde{X}_{(k+1|k+1)} &= \tilde{X}_{(k+1|k+1)} - \tilde{X}_{(k+1|k)} \\ &= \tilde{X}_{(k+1|k+1)} - A_{k+1}\tilde{X}_{(k|k)} \\ &= K_{k+1}\Phi_{k+1}(\|\omega_{(k+1|k)}\|^2)G_{k+1}^{1/2}\omega_{(k+1|k)} \end{aligned} \quad (4)$$

where $\omega_{(k+1|k)} = G_{k+1}^{-1/2}\hat{E}_{(k+1|k)}$ is one-step predicted weighted residual.

For threshold constant series $\{C_{k+1}\}$ used to control the difference between the filtered values and the predicted values of state vectors X_{k+1} , Φ_{k+1} should be suitably chosen so as to make sure that the following inequality holds

$$\begin{aligned} \|\Delta\tilde{X}_{(k+1|k+1)}\| &\leq \{\omega_{(k+1|k)}^T \omega_{(k+1|k)} \lambda_{k+1} \Phi_{k+1}^2(\|\omega_{(k+1|k)}\|^2)\}^{1/2} \\ &\leq C_{k+1} \end{aligned} \quad (5)$$

where λ_{k+1} is the maximum eigenvalue of matrix $K_{k+1}G_{k+1}K_{k+1}^T$.

It is easy to see that there are quite many different kinds of function series $\{\Phi_{k+1}\}$ satisfying inequality (5). All of the function series $\{\Phi_{k+1}\}$ are denoted as a set S :

$$S = \{\{\Phi_{k+1}\} : \Phi_{k+1} \leq 1, \sqrt{r}\Phi_{k+1}(r) \leq \frac{C_{k+1}}{\sqrt{\lambda_{k+1}}}, r \in [0, +\infty), k \in N\} \quad (6)$$

The way to select function series $\{\Phi_{k+1}\}$ from the set S is examined below.

Theorem 1 : For a linear stochastic system, when the noise series $\{x_0, \varepsilon_0, \varepsilon_1, \Lambda; \eta_1, \eta_2, \Lambda\}$ are stochastic sequences that possess a normal distribution and are mutually independent, if

the weighting matrix sequence are $G_{k+1} = H_{k+1}\Sigma_{k+1}H_{k+1}^T + R_{\eta(k+1)}$, then the function series $\{\Phi_{k+1}\} \in S$ which lead to the least errors of filtering estimation are as follows:

$$\Phi_k(r) = \begin{cases} 1, & r \leq \frac{C_k^2}{\lambda_k} \\ \frac{C_k}{\sqrt{r\lambda_k}}, & r > \frac{C_k^2}{\lambda_k} \end{cases} \quad (7)$$

where C_k is the suitably chosen threshold constant.

Proof : It follows from equations (2) and (3) that the filtering error can be expressed as follows:

$$\begin{aligned} E\|\tilde{X}_{(k+1|k+1)} - X_{k+1}\|^2 &= E\|\tilde{X}_{(k+1|k+1)} - \hat{X}_{(k+1|k+1)} + \hat{X}_{(k+1|k+1)} - X_{k+1}\|^2 \\ &= E\|\tilde{X}_{(k+1|k+1)} - \hat{X}_{(k+1|k+1)}\|^2 + E\|\hat{X}_{(k+1|k+1)} - X_{k+1}\|^2 \\ &\quad + 2E(\tilde{X}_{(k+1|k+1)} - \hat{X}_{(k+1|k+1)})^T (\hat{X}_{(k+1|k+1)} - X_{k+1}) \end{aligned}$$

Using the basic properties on conditional probability distribution, the following results may be deduced

$$\begin{aligned} E(\tilde{X}_{(k+1|k+1)} - \hat{X}_{(k+1|k+1)})^T (\hat{X}_{(k+1|k+1)} - X_{k+1}) &= E\{E(\tilde{X}_{(k+1|k+1)} - \hat{X}_{(k+1|k+1)})^T (\hat{X}_{(k+1|k+1)} - X_{k+1}) | Y_1, \Lambda, Y_{k+1}\} \\ &= E\{E(\tilde{X}_{(k+1|k+1)} - \hat{X}_{(k+1|k+1)})^T E\{\hat{X}_{(k+1|k+1)} - X_{k+1} | Y_1, \Lambda, Y_{k+1}\}\} \\ &= E\{E(\tilde{X}_{(k+1|k+1)} - \hat{X}_{(k+1|k+1)})^T \{E\hat{X}_{(k+1|k+1)} - E(X_{k+1} | Y_1, \Lambda, Y_{k+1})\}\} \\ &= 0 \end{aligned}$$

Note that the project property of Kalman filtering has been used in the last equation of the above expression. So the following result can be obtained:

$$\begin{aligned} E\|\tilde{X}_{(k+1|k+1)} - X_{k+1}\|^2 &= E\|\tilde{X}_{(k+1|k+1)} - \hat{X}_{(k+1|k+1)}\|^2 + E\|\hat{X}_{(k+1|k+1)} - X_{k+1}\|^2 \end{aligned} \quad (8)$$

In order to investigate how to minimize the error in equation (8), equations (2) and (3) are used to deduce the following expression:

$$E\|\tilde{X}_{(k+1|k+1)} - \hat{X}_{(k+1|k+1)}\|^2 = E\{[1 - \Phi_{k+1}(r_{k+1})]^2 \|K_{k+1}\hat{E}_{(k+1|k)}\|^2\} \quad (9)$$

It is obvious that function series $\{\Phi_k\} \in S$ which were prescribed in the equation (7) minimize the equation (9). ■

In the following part of this section, we will discuss how to calculate weigh matrix sequence $\{G_{k+1}\}$ and threshold constants C_{k+1} ($k = 1, 2, \Lambda$).

Hypothesis 1. Suppose that $\{\varepsilon_k\}$ and $\{\eta_k\}$ are two independent noise sequences and they satisfy following relationships:

- a). $E\varepsilon_k = 0, E\eta_k = 0, Ex_0 = \bar{x}_0, Ex_0x_0^T = R_0 + \bar{x}_0\bar{x}_0^T$;
- b). $E\varepsilon_k\varepsilon_k^T = R_{\varepsilon(k)}\delta(k-k), E\eta_k\eta_k^T = R_{\eta(k)}\delta(k-k)$;
- c). $E\varepsilon_k\eta_k^T = 0, Ex_0\varepsilon_k^T = 0, Ex_0\eta_k^T = 0$

Hypothesis 2 : Suppose that $\{\varepsilon_k\}$ and $\{\eta_k\}$ are two independent noise sequences and they satisfy following distributions:

$$x_0 \sim N(\bar{x}_0, R_0); \varepsilon_i \sim N(0, R_{\varepsilon(i)}); \eta_i \sim N(0, R_{\eta(i)})$$

Lemma 1 : If a linear stochastic system defined by equation (1) satisfies hypothesis 1 and hypothesis 2, then stochastic sequence $\{x_0, \varepsilon_0, \Lambda, \varepsilon_k; \eta_1, \Lambda, \eta_k\}$ obeys the law of a multivariate normal distribution and their joint covariance matrix C_a is as follows:

$$C_{a(k)} = \begin{pmatrix} C_{a1}(k) & 0 \\ 0 & C_{a2}(k) \end{pmatrix} = C_{a1}(k) \& C_{a2}(k) \quad (10)$$

where operator “&” represents the block-diagonal matrix formed by the two block matrices on either side of the operator; $C_{a2}(k)$ is the covariance matrix of the measured noise sequence, which can be expressed as the inverse matrix of the tri-diagonal matrix $A = (a_{i,j})_{k \times k}$ which is defined as follows

$$\begin{cases} a_{1,1} = R_0^{-1} \& A_1^T R_{\varepsilon(1)}^{-1} A_1 \\ a_{i,i} = R_{\varepsilon(i-1)}^{-1} \& A_i^T R_{\varepsilon(i)}^{-1} A_i \\ a_{k,k} = R_{\varepsilon(k-1)}^{-1} \\ a_{i,i+1} = A_i^T R_{\varepsilon(i)}^{-1} \\ a_{i-1,i} = R_{\varepsilon(i)}^{-1} A_i \\ a_{i,j} = 0 \quad (|i-j| \geq 2, i, j = 1, \dots, k) \end{cases} \quad (11)$$

In formula (11) matrix $C_{a2}(k)$ is the covariance matrix of the measured noise sequence $\{\eta_k\}$.

$$C_{a2}(k) = \text{diag}\{R_{\eta(1)}, \Lambda, R_{\eta(k)}\}$$

Lemma 2 : If the linear stochastic system defined by equation (1) satisfies hypothesis 1 and hypothesis 2, the joint distributions of sequence $\{x_0, x_1, \Lambda, x_k; y_1, \Lambda, y_k\}$ are normal and their covariance matrix is given by the following relationships

$$C_b = \begin{bmatrix} C_{a1} & M & C_{a1}H^T \\ \Lambda & \Lambda & \Lambda \\ HC_{a1} & M & HC_{a1}H^T + C_{a1} \end{bmatrix}, H = \begin{bmatrix} 0 & H_1 & \Lambda & 0 \\ M & M & M \\ 0 & 0 & \Lambda & H_k \end{bmatrix}$$

where the superscript τ denotes transpose of a matrix.

Analyzing these two lemmas described above, it is found that the joint distribution of sequence $\{y_1, \Lambda, y_k\}$ is also normal.

If the linear stochastic system defined by equation (1) satisfies those two hypotheses, denoting the joint covariance matrix of $\{y_1, \Lambda, y_k\}$ by $D_{(k)}$ and the covariance matrix of $\hat{E}_{(k+1|k)}$ by Ω_{k+1} , it is easy to verify the properties of Kalman filtering algorithms that

$$\Omega_{k+1} = d_{k+1} - D_{d(k)}^{\tau} D_{(k)}^{-1} D_{d(k)} \quad (13)$$

where $d_{k+1} = \text{cov}(y_{k+1}, y_{k+1})$, the auto-variance matrix of stochastic vector y_{k+1} ; matrix $D_{d(k)}$ is the covariance matrix of $\{y_1, \Lambda, y_k\}$ and y_{k+1} .

Weighting matrix G_{k+1} and weighted residual $\omega_{(k+1|k)}$ can be calculated as follows

$$\begin{cases} G_{k+1} = \Omega_{k+1} = d_{k+1} - D_{d(k)}^{\tau} D_{(k)}^{-1} D_{d(k)} \\ \omega_{(k+1|k)} = \Omega_{k+1}^{-1/2} \hat{E}_{(k+1|k)} \end{cases} \quad (14)$$

It can be proven that $\omega_{(k+1|k)}$ follows the standard multivariate normal distribution. Supposing that the measured information is an m -dimensional vector, the weighted residual $\omega_{(k+1|k)}$ is an m -dimensional standard normal variable, the norm of which is equal to r_{k+1} :

$$r_{k+1} = \hat{E}_{(k+1|k)}^{\tau} \Omega_{k+1}^{-1} \hat{E}_{(k+1|k)} \sim \chi^2(m) \quad (15)$$

where $\chi^2(m)$ denotes the χ -distribution of degrees-of-freedom m .

Since the random variable r_{k+1} satisfies the distribution $\chi^2(m)$, the threshold constant $c^{\alpha}(m)$ satisfies

$$P(r_{k+1} > c^{\alpha}(m)) = \alpha$$

Generally parameter α is taken as 0.05 or 0.025. If $r_{k+1} \leq c^{\alpha}(m)$, it is believed with $(1-\alpha) \times 100\%$ confidence that the additional information from the measured data y_{k+1} is reasonable; otherwise, when $r_{k+1} \geq c^{\alpha}(m)$, it is also believed with $\alpha \times 100\%$ confidence that the additional information from the measured data y_{k+1} is unreasonable and hence this additional information must be discarded.

From the above analysis and considering equation (6), it is known that C_{k+1} can be given by

$$C_{k+1} = (\lambda_{k+1} c^{\alpha}(m))^{1/2} \quad (16)$$

where λ_{k+1} is the maximum eigenvalue of matrix $K_{k+1} K_{k+1}^{\tau}$.

In specific applications in engineering, C_{k+1} may be chosen with our experience.

V. NUMERICAL SIMULATION

Supposing that the coefficient matrices and covariance matrix of errors defined by equation (1) are as follows

$$A_k = \begin{bmatrix} 1.0 & 0 & 0 \\ 0.2 & 1.0 & 0 \\ 0 & 0.3 & 1.0 \end{bmatrix}, \quad H_k = \begin{bmatrix} 0.9 & 0 \\ 0 & 0.2 \\ 0 & 0.7 \end{bmatrix}^{\tau}$$

and the covariance of the model errors are respectively equal to

$$\begin{cases} R_{\delta(k)} = \text{diag}\{0.5, 0.5, 0.5\} \\ R_{\eta(k)} = \text{diag}\{0.1, 0.1\} \end{cases}$$

Using Monte Carlo method and selecting the initial states of the system as the following

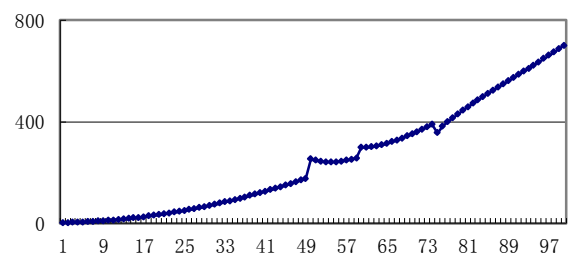
$$\begin{cases} x_{1(0|0)} = 1.3 \\ x_{2(0|0)} = 1.5 \\ x_{3(0|0)} = 2.3 \end{cases}$$

$$\Sigma_{(0|0)} = \begin{bmatrix} 0.01 & & \\ & 0.01 & \\ & & 0.01 \end{bmatrix}$$

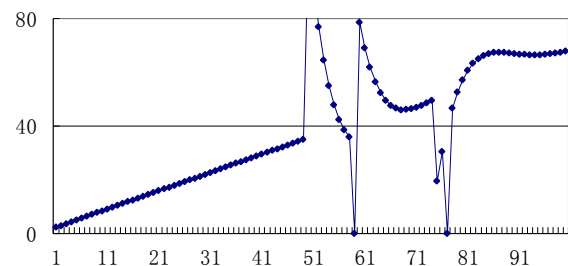
one hundred pieces of simulation data are generated and denoted by the set S . Let deviations of the 50th and 75th pieces of data be

$$\Delta y_1 = (-1)^i 100, \Delta y_2 = (-1)^{i+1} 5 \quad (i = 50, 75)$$

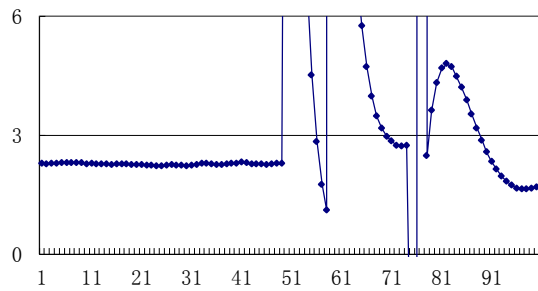
The new set with these two outliers is denoted by S^* . The estimates on S^* made by equations (1) and (2) are shown in figures 1 and 2.



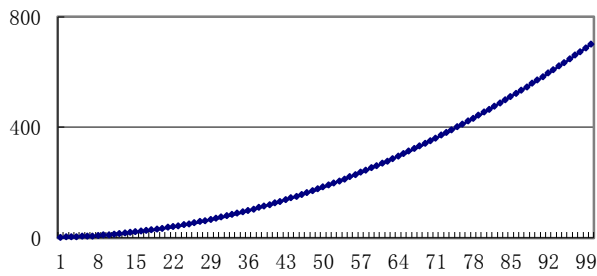
(a) Filtering Estimation of X_1



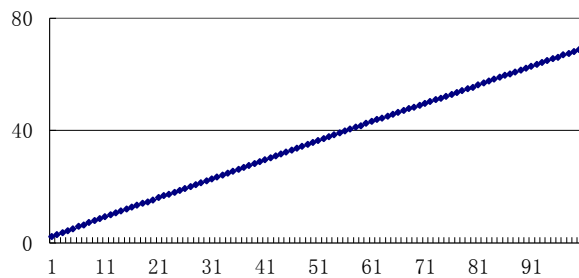
(b) Filtering Estimation of X_2



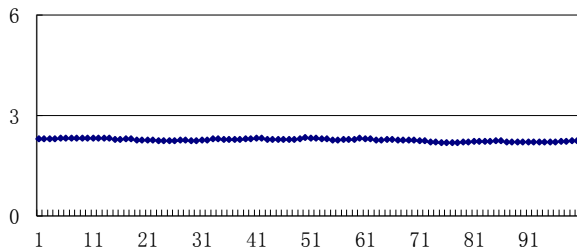
(c) Filtering Estimation of X_3
Figure 1 Kalman Filters of State Variables $X \in R^3$



(a) Filtering Estimation of X_1



(b) Filtering Estimation of X_2



(c) Filtering Estimation of X_3

Figure 2 Fault-tolerant Filters of State Variables $X \in R^3$

It can be seen clearly that the outliers have a very negative effect on conventional Kalman filtering and the outliers-tolerant modification method proposed in the paper is capable of overcoming this negative effect and is reliable.

ACKNOWLEDGEMENTS

The research was supported by the Sweden Institute Grant (SI 05483/2005-210), the National Natural Science Foundation of China (No.61074077) and the Tianyuan Fund of National Natural Science Foundation of China (No.11026224).

The first author would like to thank Dean Ingrid Melinder of the Computer Science and Communication (CSC) of the Royal Institute of Technology (KTH) for her friendly help when he visited the CSC of KTH of Sweden.

REFERENCES

- [1] Hu Shaolin, Fan Jingcheng. "Bounded Influence Filter of the Dynamic-Measurement System". Control Theory and Applications, 1993.10(1): 36-45
- [2] Han Congzhao, Wang Yuejuan, Wan Baiwu. Stochastic System Theory. Press of the Xi'an Jiaotong University, 1987
- [3] Martin D, Mintz. "Robust Filtering and Prediction for Linear System with Uncertain Dynamics — a game theoretic approach". IEEE Trans. Automatic Control, 1983, AC-18
- [4] Martin D. Robust Method for Time Series, Academic Press Inc, 1980
- [5] Hu Shaolin, Wang Xiaofeng, Karl Meinke, Huajiang Ouyang. Outlier-tolerant fitting and online diagnosis of outliers in dynamic process sampling data series. in Artificial Intelligence and Computational Intelligence, pp.195-204, LNAI 7004 (Deng Miao, Lei Wang, eds) Springer Press, 2011
- [6] F Carlos, S Alcalá, J Qin. Unified Analysis of Diagnosis Methods for Process Monitoring. In: Proc of the 7th IFAC Symp on Fault Detection, Supervision and Safety of Technical Processes, Spain, 2009, 1007-1012
- [7] Nihal Y, Bovas A, MacGregor J. A kalman filter in the presence of outliers Communications in Statistics - Theory and Methods Volume 20, Issue 5-6, 1991
- [8] Durovic, Z.M., Kovacevic, B.D.: Robust estimation with unknown noise statistics. IEEE Transactions on Automatic Control 44, 1292-1296, 1999
- [9] Chan, S.C., Zhang, Z.G., Tse, K.W.: A new robust Kalman filter algorithm under outliers and system uncertainties. In: IEEE International Symposium on Circuits and Systems. IEEE, 4317-4320, 2005
- [10] Jo-Anne Ting, Evangelos Theodorou, Stefan Schaal. A Kalman Filter for Robust Outlier Detection. Computational Learning & Motor Control Lab University of Southern California, IROS, 2007

Handsets Malware Threats and Facing Techniques

Marwa M. A. Elfattah

Computer Science Dep,
Faculty of Computers and Information,
Helwan University, Cairo, Egypt

Aliaa A.A Youssif

Computer Science Dep,
Faculty of Computers and Information,
Helwan University, Cairo, Egypt

Ebada Sarhan Ahmed

Computer Science Dep,
Faculty of Computers and Information,
Helwan University, Cairo, Egypt

Abstract—Nowadays, mobile handsets combine the functionality of mobile phones and PDAs. Unfortunately, mobile handsets development process has been driven by market demand, focusing on new features and neglecting security. So, it is imperative to study the existing challenges that facing the mobile handsets threat containment process, and the different techniques and methodologies that used to face those challenges and contain the mobile handsets malwares. This paper also presents a new approach to group the different malware containment systems according to their typologies.

Keywords – mobile; malware; security; malicious programs.

I. INTRODUCTION

Recently, mobile handsets are becoming more intelligent and complex in functionality, much like PCs. Moreover, mobiles are more popular than PCs, and are being used more and more often to do business, access the Internet, access bank accounts, and pay for goods and services. This resulted in an increased number of criminals who wants to exploit these actions for illegal gains.

Today's malware is capable of doing many things, such as: stealing and transmitting the contact list and other data, locking the device completely, giving remote access to criminals, sending SMS and MMS messages etc. Mobile malware causes serious public concern as the population of mobile phones is much larger than the population of PCs [1-6].

The first proof-of-concept of mobile malware was "Cabir" [7], which was proposed in 2004 targeting Symbian OS. After that, mobile malware evolved rapidly during the first two years (2004 - 2006) of its existence. A wide range of malicious programs targeting mobile phones appeared, and these programs were very similar to malware which targeted computers: viruses, worms, and Trojans, the latter including spyware, backdoors, and adware. Now a days , the amount of malware for mobile devices have been duplicated more than one time. This shows that the growth rate demonstrated between 2004 and 2006 has been maintained.

In response to this increasing threat, this paper was developed as a survey to contain this critical phenomenon. The paper is organized to introduce the challenges of the mobile handsets malware containment and facing operation in section 2. Section 3 shows the different techniques which were proposed by researchers to face the mobile malware. Then, in section 4, a new approach to group the malware containment systems according to their typologies is introduced. To the best of authors' knowledge, this approach

of grouping is never introduced before. Finally, in section 5, this work is concluded.

II. MOBILE MALWARE FACING CHALLENGES

Although the great dangerous of mobile handset malware, and the importance of finding a solution to limit this danger, the task of facing mobile handset malware and limiting their harm has a lot of obstacles and is not easy to be faced. In this section most of those obstacles are concluded [1-6]:

- Some of mobile handset users treat mobile handset malware as a problem which has not happened yet, or believe that it's not an issue which really concerns them.
- A mobile handset has limited processing power and storage capacity, unlike resource-rich PCs, a detection framework should not consume too much of the device resources, including CPU, memory, and battery power, the overhead for executing the detection framework should be kept to a minimum.
- Most new malicious programs for mobile handset devices are hybrids, containing functionality from two or more different types of malware.
- When virus writers realized that there was no clear leading operating system for mobile devices, they also realized it wouldn't be possible to target the majority of mobile device users with a single attack. Because of this, they started focusing less on writing malware which targeted specific platforms, and more on creating programs capable of infecting several platforms.
- While a computer is primarily connected to the internet via IP networks, a mobile handset also connects to the cellular network through SMS/MMS services, as well as its Bluetooth interface that is frequently used to interact with other devices. These interfaces are quickly becoming the new infection vector for viruses, which makes the mobile handset susceptible to get infected even when it is disconnected from the internet.
- A mobile handset is highly mobile and always on, resulting in a greater degree of difficulty in quarantining the virus in a local region.
- To evade detection, malware writers are increasingly using polymorphic coding techniques. Polymorphism is a process through which malicious code modifies

its appearance to evade detection without actually changing its underlying functionality. These techniques include everything from modifying the names of internal variables and subroutines, changing the order in which instructions appear in the body of malware, to encrypting most of the malware code, only leaving in the clear text the instructions necessary to decrypt the code [1]. In addition to changing the appearance of malware via polymorphism, new malware can further change their behavior, going through metamorphism; metamorphic code actually changes the functionality of malware, while hiding its payload using obfuscation and encryption [1]. When metamorphic techniques are used in conjunction with polymorphism, malware of this kind are much harder to detect, analyze, and filter.

III. MOBILE MALWARE FACING TECHNIQUES

An effective detection frame work should be able to detect diverse types of malware and malware variants, keeping both false-negatives and false-positives below a certain acceptable threshold; also it should not consume the device resources.

There are set of approaches for preventing mobile handset from malware as shown in Fig 1. The simplest of them is to only trust and install digitally signed applications [4]. This ensures that the software has undergone a standard testing procedure as part of being signed. However, given the vast number of mobile applications available on the Internet, especially peer-to-peer sites, one cannot expect all applications to be signed with a certificate. An application that has been self-signed cannot be trusted to be free of malicious code. Moreover, even when an application is signed by a trusted CA, a malicious program can still enter the system via downloads (e.g., SMS/MMS messages with multimedia attachments), and it may exploit known vulnerabilities of an unsigned helper application.

Signature-based detection is another well known procedure for handling mobile handset malwares. It relies on static file signature which make it vulnerable to simple obfuscation, polymorphism, and packing techniques. Also the need of a huge database to store a signature for each malware makes this technique unsuitable for mobile devices which suffer from limited resource.

An alternative to signature-based methods is the behavioral-based detection, which has been emerged as a promising way of preventing the intrusion of spyware, viruses and worms.

A. Signature-Based Detection

Signature-based detection is one of the most known techniques for malware detection. To identify malwares,

signature-based detection systems compare the contents of a file to a dictionary of malware signatures. There are well developed signature-based techniques for malware detection on the PC domain, but it requires considerable effort to adapt these techniques for mobile handsets. Also, signature-based detection techniques are unsuitable for mobiles because those techniques require a new signature for every single malware variant. However, mobile handsets have severe resource constraints in terms of memory and power.

Some of researchers attempted deal with these difficulties, and to adjust signature-based detection algorithm to fit the mobile device. Deepak and Guoning [8, 9] suggested a system for detecting malware using malware signatures. This system automatically extracts a set of signatures from existing malware samples. It reduced the number of signatures by using a common signature for a malware and its variants.

Also, it minimized the total false alarm rate of malware detection by extracting signatures that are most uncommon within mobile network traffic. Deepak and Guoning outlined the considerations for malware detection on mobile devices and proposed a signature matching algorithm with low memory requirements and high scanning speed. They used hash table and sub-signature matching to scans the network traffic.

Although deepak solution consumes less than 50% of the memory used by Clam-AV while maintaining a fast scanning rate [8], signature-based detection methods still suffer from a lot of other weaknesses which make most of proposed signature-based solutions unsuitable for mobile handset [1, 4, 6]. Signatures are created using static information, thus being vulnerable to simple obfuscation, polymorphism, and packing techniques. Also, it is challenging to distribute virus signatures files to the mobile handsets in a timely manner. Even though, only one signature is required for a malware and its variants, the amount of those signatures is still more than the amount of malware behavior signatures, which are changed rarely.

Moreover, anti-virus solutions for mobile devices rely only on signature-based detection could not be considered as Conclusive solutions. Although the infected files are deleted by the anti-virus tool, the underlying vulnerability is not patched. As a result, a cleaned handset may get infected again by another instance of the same virus, requiring repeated cleanup.

B. Behavioral -Based Detection:

In behavioral-based detection techniques, the behavior of an application is monitored and compared against a set of malicious and/or normal behavior profiles. The malicious behavior profiles can be specified as global rules that are applied to all applications, as well as fine-grained application-specific rules [4-10].

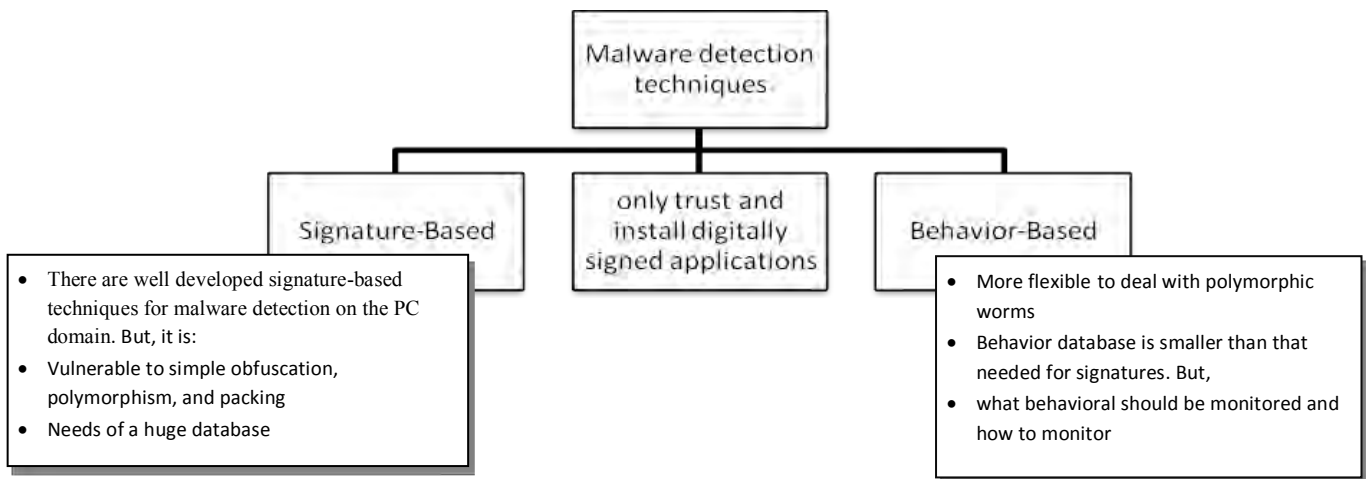


Figure 1: Mobile malware detection techniques

Behavioral detection is more flexible to deal with polymorphic worms or code obfuscation, because it assesses the effects of an application based on more than just specific payload signatures. Considering the fact that a new malware variant is usually created by adding new functionality to existing malware or modifying obsolete modules with fresh ones, this abstraction is effective for detecting previously-unknown malware variants that share a common behavior exhibited by previously-known malware. A typical database of behavior profiles and rules should be smaller than that needed for storing specific payload signatures of many different classes of malware. This makes behavioral detection methods particularly suitable for handsets.

One common problem with behavioral detection, however, is specification of what constitutes normal or malicious behavior that covers a wide range of applications [4], while keeping false positives (incorrect identification of a benign activity as malicious) and false-negatives (failure to identify malicious activities) low. Another one is the reconstruction method of potentially suspicious behavior from the applications, so that the observed signatures can be matched against a database of normal and malicious signatures. Heuristics on what behavioral should be monitored and how to monitor and collect behavioral vary. In the following, some of researchers' efforts to produce methodologies for malware detection and relevant behavioral measurements are concluded.

Code Analysis: There are several methods to analysis the executable code. For example, D. Venugopal et-el [11] proposed an algorithm to monitor each application and library (e.g., a dynamic link library (DLL)) that a process attempts to load. This information is then compared against lists of authorized and unauthorized applications and libraries.

They observed that most viruses in the mobile domain demonstrate common functionalities, such as, deleting system files and sending MMS messages, so mobile viruses are classified into different families or classes based on their functionalities. All virus variants in a family share a common malicious core behavior. Each malware needs to

use certain dynamic link libraries (DLLs) to implement their functions. The DLL functions used by a virus give a good insight into the functionality of this virus. Therefore, the imported DLL functions were used as features for virus detection. They considered their method is computationally efficient since the DLL functions are easy to be extracted from the executable files [11].

Carsten Willems et-al developed CWSandbox [12], which is a malware analysis tool that employ dynamic malware analysis, API hooking and dynamic linked library (DLL) injection techniques to implement the necessary rootkit functionality to avoid detection by the malware. Using dynamic analysis techniques, they observed malware behavior and analyzes its properties by executing the malware in the sandbox. The analysis process can be done by taking an image of the complete system state before malware execution and comparing it to the complete system state after execution, or by monitoring the malwares actions during execution with the help of a specialized tool, such as a debugger.

To observe a given malware sample's control flow, it is important to access the application API functions. One possible way to achieve this is by hooking - intercepting a call to a function. When an application calls a function, it's rerouted to a different location where customized hook function resides. The hook then performs its own operations and transfers control back to the original API function or prevents its execution completely.

DLL code injection help in implementing of API hooking in a modular and reusable way. However, API hooking with inline code overwriting makes it necessary to patch the application after it has been loaded into memory. To be successful, the hook functions must be copied into the target application's address space so they can be called from within the target - this is the actual code injection - and bootstrap the API hooks in the target application address space using a specialized thread in the malware memory.

CWSandbox system outputs a behavior-based analysis that is; it executes the malware binary in a controlled environment so that it can observe all relevant function calls

to the system API, and generates a high-level summarized report from the monitored API calls. To enable fast automated analysis, the CWSandbox is executed in a virtual environment so that the system can easily return to a clean state after completing the analysis process. But, this approach suffers from some drawbacks, such as, slower execution and device overloading.

Also, Liang Xie et-el [13] assumed that malware always launch attacks from the application software. From the application's point of view, malware attacks always cause anomalies in process states and state transitions. Such anomalies are reflected through malware function (API) calls, usages of system resources, and requests for system services. So, they adopt function call-trace techniques and human intelligence techniques in the context of cell phones to identify process misbehavior.

Also they noted that, each cell phone user has his own unique and private operational patterns (e.g., while operating keypad or touch-screen), which cannot be easily learned and simulated by malware. From these two aspects, and their behavior-based malware detection system (pBMDS) provides comprehensive protection against malware. pBMDS leverages a Hidden Markov Model (HMM) to learn process behaviors (states and state transitions) and additionally user operational patterns, such that it can effectively identify behavior difference between malware and human users for various cell phone applications.

File system Monitoring: A file system can be monitored through a number of aspects such as checking file integrity, file attributes, or file access attempts. Both file integrity and attribute checking can only be determined if a change has taken place [1], but file access attempts can be predetermined.

In checking for file access attempts, X. Zhang, et-al [14, 15] proposed a mandatory access control (MAC) system to strictly controls - according to some predefined sets of rules - the interactions between subjects (e.g., services or processes) and objects (e.g., files, sockets, etc.), which are differentiated based on the labels assigned to them. In the system an agent with a shim - a layer of code placed in between existing layers of code - can monitor all attempts to access critical files and stop suspicious attempts by comparing policies with the characteristics of the current attempt, such as which user / application attempts to access what file with a particular type of access (i.e., read, write, or execute).

The main advantage of this approach on mobile handsets is that, kernel-level mechanisms are intrinsically trusted, simply because that the kernel is a part of the trusted computing base. Also a MAC-based isolation is better than virtualization techniques due to the pure performance. Since mobile phones have limited computational capabilities and low power consumption requirements, virtualization becomes an impractical solution.

However, Although MAC mechanisms consume substantial computing power on PC platforms (due to vast

number of subjects and objects), mobile handsets in contrast are still limited and cannot be compared to classical PC environments in this regard. This significantly simplifies the security policies and improves the potential performance of MAC mechanisms on mobile devices. But kernel-level solutions are too difficult to be implemented.

Power Consumption Monitoring: While most malicious code attacks on handhelds aim to damage software resources, intentional abuse of hardware resources (e.g., CPU, memory, battery power) has become an important, increasing threat. In particular, malware targeting the burning/depletion of battery power are extremely difficult to be detected and prevented, mainly because users are usually unable to recognize this type of anomaly on their handhelds and the battery can be deliberately and rapidly drained in a number of different ways.

H. Kim et-el [1] have designed a malware-detection framework, which is composed of a power monitor and a data analyzer. The former collects power samples and builds a power consumption history with the collected samples, and the latter generates a power signature from the power consumption history. The data analyzer then detects an anomaly by comparing the generated power signature with those in a database.

VirusMeter is another system that was developed by Liu L et-el [16] who illustrated that, by monitoring power consumption on a mobile device, VirusMeter catches misbehaviors that lead to abnormal power consumption. VirusMeter relies on a concise user-centric power model that characterizes power consumption of common user behaviors based on the number or the duration of the user actions, such as, the duration of Call, the number of SMS, and etc.

These works have shown that power anomaly is an effective indicator for suspicious activities on mobile phones. To identify the causes of these activities is still a challenge for power-based malware detection as the power consumption for normal behavior is yet to be accurately quantified. Another challenge is that existing mobile handsets is not able to provide sufficient precision for power consumption measurement without involving extra measuring devices like an oscilloscope [10].

Communicational Statistical Modeling: Statistical modeling for malware is usually used as a collaboration defense for preventing the malware spreading over the network. D. Venugopal, Hu. Guoning designed SmartSiren [2] which aims to detect worms exploiting SMS messaging and Bluetooth communication. This system keeps track of the communication activities on the device. In cases where abnormal activities have been locally identified, alerts are sent to both infected devices and a subset of the uninfected devices, which may be in contact with an infected device, based on the users' contact lists and mobility profiles.

IV. THE PROPOSED GROUPING APPROACH FOR MALWARE FACING METHODOLOGIES BASED ON TYPOLOGIES

To the best of authors' knowledge, there is not any study that groups the malware containment systems according to

their typologies. But, it is widely found that, the done work on the field of mobile malware detection and prevention can be grouped into three complementary typologies, as shown in Fig 2.

A. Device-Based Detection Typology:

Due to the danger of the malware attack which aim the mobile handsets, a lot of researchers have concentrate all of their attention on the device based solutions. They tried to face the malware attacks by proposing detection and prevention systems that are completely built on the device and they never affect or are affected by the infrastructure. For example, Aciimez et-al [14] developed kernel level but general-purpose mandatory access control (MAC) mechanisms for main stream operating systems. Typically, a MAC system strictly controls the interactions between subjects (e.g., services or processes) and objects (e.g., files, sockets, etc.), which are differentiated based on the labels assigned to them.

Also, G Tuvell et-el[17] developed a system and method for detecting malware with in a device by modeling the behavior of malware and comparing a suspect executable with the model. The system and method extracts feature elements from malware-infected applications. Using malware-free and malware-infected applications as training data, the system and method heuristically trains the rules and creates a probability model for identifying malware. To detect malware, the system and method scans the suspect executable for feature sets and applies the results to the probability model to determine the probability that the suspect executable is malware-infected.

B. Infrastructure-Based Detection Typology:

Another set of researchers noted that, each device affects and is affected by a not negletable set of other devices, and the malware facing will be more proactive if it was in a collective manner instead of other individual solutions. Those researchers preferred to propose solutions that concern with the complete infrastructure. Their solutions are based on collecting information from the infrastructure components in organized manner. The collected information is used in the protection and the malware detection solutions.

For example, V. Karyotis et-el [18] have study the propagation of malware over a wireless ad hoc network. They proposed a probabilistic model that is able to model and capture the aggregated behavior of a large ad hoc network attacked by a malicious node, where legitimate

network nodes are prone to propagate infections they receive to their neighbors. They used the Norton equivalent representation of the proposed network model that allowed them to acquire analytical results of the behavior of the system in its steady state. Depending on the acquired relations, they were able to identify the critical system parameters and the way they affect the operation of the network.

In order to analyze the influence of various system parameters on the network operation and identify which of them can be exploited by an attacker or by the network itself, they focused on the average number of infected nodes and the average throughput of the non-infected nodes, which could be indicative of the overall asymptotic system behavior independently of a specific network instant.

Furthermore, the Infection efficiency of an attack was obtained through simulation and used as a comprehensive attack evaluation metric in order to evaluate the impact of attackers on the network for a specific time period and scenario, indicating potential short-term variations and effects. Also, some insight regarding the behavior and evolution of the system when multiple attackers operate simultaneously and independently in the network is gained via modeling and simulation.

SmartSiren [2] is another example of infrastructure-based solution to securing mobile phones despite the limitations of post-infected detection. The goal of SmartSiren is to halt the potential virus outbreak by minimizing the number of mobile handsets that will be infected by a new released virus. The outbreak of viruses must affect many mobile handsets and cause noticeable changes in their behavior. Thus, early detection of viruses can be achieved by keeping track of the device activities even in a coarse granularity.

In this system, each mobile handset runs a light-weight agent, while a centralized proxy is used to assist the virus detection and alert processes. Each mobile handset agent keeps track of the communication activities on the device, and periodically reports a summary of these activities to the proxy. In cases where abnormal activities have been locally identified, a mobile handset may also submit a report immediately to the proxy. On the other hand, the proxy detects any single-device or system-wide viral behaviors. When a potential virus is detected, the proxy sends targeted alerts to both infected devices and a subset of the uninfected devices, which may be indirect contact with an infected device, based on the users' contact lists and mobility profiles.

For each user, based on the user-submitted communication log, the proxy would keep track of the average number of communications that each user initiates each day using a 7 days moving average window. The summation of the 7 days moving average captures the normal usage of each user and is considered as a threshold. In addition, each day the mobile handset user's agent will count the number of communication that the users have initiated. When the user's daily usage exceeds the threshold, the user would be moved from normal state into over-usage

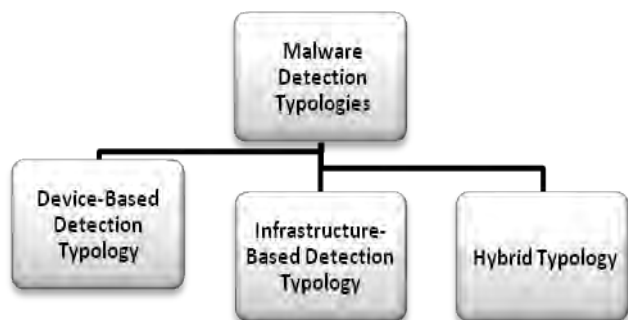


Figure 2: Malware Detection Typologies
The proposed grouping approach for malware facing methodologies based on typologies

state. The over-usage state does not guarantee that a particular handset is infected. The proxy also monitors how many users would exceed their threshold. When this daily count exceeds wildly from the average, it can suggest that an aggressive viral outbreak has occurred.

Also, Bose and Shin [3, 4, 6] discussed an agent-based malware modeling (AMM) framework that to investigate malware exploiting SMS/MMS and Bluetooth vulnerabilities on cellular handsets. In AMM, a mobile network was modeled as a collection of autonomous decision-making entities called agents. The agents represent networked devices within the network such as PDAs, mobile phones, service centers and gateways. In case of agents representing mobile devices, the connectivity changes as users roam about the physical space of the network. The behaviors of the agents are specified by a set of services running on them.

Thus, there are two types of topologies in Bose and Shin simulation environment. The physical connectivity, whereas the logical connectivity is determined by the messages exchanged among the agents. An agent may participate in multiple logical topologies corresponding to different services like email, IM, SMS, etc. They also group the agents in a hierarchical manner. For example, agents representing cellular base stations can keep track of mobile devices in their respective cells. Accordingly, these agents are able to collect information aggregated over the individual devices in their respective cells. This capability of higher-level agents to aggregate observations collected from lower-level agents reflects real-life processing of information within a mobile network. The information processed at these different levels can also be used to activate different response mechanisms against a spreading malware.

C. Hybrid Typology:

As it is illustrated previously, mobile handsets have only limited resources in terms of computation, storage, battery power. Also mobile handset users always annoy from any slowing down on the mobile performance, so, the user should never be disturbed with the existence of the detection systems. Those limitations restrict the device base solutions and harden their rule. Although Infrastructure-based solutions are computationally more expensive than device-based solutions, they offload most of the processing burden from the resource-constrained mobile handsets, thus minimizing the performance penalty on the mobile handsets.

Also, it is important, for accurate virus detection and prompt alerts, the mobile handsets must collaborate with each other. Infrastructure-based solutions simplify this collaboration among the mobile handsets by using a centralized agent that serves such collaboration. Of course, the centralized agent marked itself as a performance bottleneck and a single point of failure in the system. To improve the scalability and resiliency, one can extend the architecture with multiple agents.

Some of researchers have proposed a hybrid approach for malware detection and prevention. On this approach,

part of the detection system is placed in the device, and the other part is placed on the infrastructure. For example, A. Schmidt et-al [19] introduces an approach of how to monitor mobile handsets in order to extract values that can be used for remote anomaly detection. Therefore, it has to be learned what is the normal behavior of a user and device in order to be able to distinguish between normal and abnormal, possibly malicious actions. The extracted features are sent as vector to a remote system, taking the responsibility for extended security measures away from the probably unaware user. These vectors can be used for methods from the field of artificial intelligence in order to detect abnormal behavior.

Another framework was proposed by H. Kim et-al [1], which was composed of a power monitor and a data analyzer. The former on the device collects power samples and builds a power consumption history with the collected samples, and the latter on the remote server generates builds a power signature from the power consumption history. The data analyzer then detects an anomaly by comparing the generated power signature with those in a database.

Also, Michael Becher [20] has said that, a promising approach is an automatic dynamic analysis, where system calls are logged and afterwards analyzed for malicious behavior. Because of mobile handset limitations, this cannot be done efficiently on the mobile handset. They designed a Mobile Sandbox to analysis the collected samples in a mobile dynamic malware analysis system. It executes the sample in an environment (the sandbox), where it can watch the steps of the investigated sample. An important requirement to ensure the integrity of the analysis is logging to a remote place rather than saving the log on the device only. Mobile Sandbox implements this communication of the device to the host system with a TCP connection over ActiveSync.

V. CONCLUSION

Due to their flexible communication and computation capabilities, and their resource constraints, mobile handsets are glued victim to malwares. A mobile handset can be attacked from the Internet since mobile are Internet endpoints, or it can be Infected from compromised PC during data synchronization; also it can have a peer mobile attack or infection through SMS/MMS and Bluetooth.

Although the difficulties of building a malware detection systems, some of researchers have concerned with this field of research. Set of researcher have done work to adjust the existed PC's signature-based detection systems to be suitable for mobiles. But, signature-based solutions have a lot of weakness, so another set of researchers preferred behavioral based solutions due to their flexibility to deal with polymorphic worms, and to the small amount of data needed to be stored. However, on behavioral based solutions, it is important to specify what constitutes normal or malicious behavior that covers a wide range of applications. This paper presented and analyzed some of researchers' effort on that aspect

On other hand, according to detection system typologies, this paper has grouped the detection systems into three complementary typologies, which are device-based detection typology, infrastructure-based detection typology. Device-based detection typology faces the resource limitation constrains, and has no way that facilitate the communication and the collaboration between detection systems. However infrastructure-based detection typology is computationally more expensive and the centralized agent marked itself as a performance bottleneck and a single point of failure in the system. In hybrid topologies, a part of the detection system is placed in the device, and the other part is placed on the infrastructure.

REFERENCES

- [1] H. Kim, J. Smith, G. Shin, "Detecting energy-greedy anomalies and mobile malware variants", The International Conference on Mobile Systems, Applications, and Communications (MobiSys), ACM/USENIX, pp. 239–25, 2008.
- [2] J. Cheng, S. Wong, H. Yang, Lu. Songwu, "Smartsiren: virus detection and alert for smartphones", The International Conference on Mobile Systems, Applications, and Communications , MobiSys, pp. 258-271, ACM, Jun. 2007.
- [3] A. Bose, G. Shin., "On mobile viruses exploiting messaging and bluetooth services", International Conference on Security and Privacy in Communication Networks ,SecureComm, IEEE, PP. 1-10, Aug. 2006.
- [4] A. Bose, "Propagation, detection and containment of mobile malware", PhD thesis , the university of Michigan, 2008.
- [5] G. Chuanxiong, J. Wang, Z. Wenwu, "Smart-phone attacks and defenses", Third Workshop on Hot Topics in Networks, HotNets III, San Diego, CA, 2004.
- [6] A. Bose, X. Hu Kang G. Shin, T. Park, "Behavioral detection of malware on mobile handsets", International Conference on Mobile Systems, Applications, and Communications , MobiSys08, pp. 225-238, June, 2008.
- [7] <http://www.viruslist.com/en/analysis>.
- [8] D. Venugopal, Hu. Guoning, "Efficient signature based malware detection on mobile devices", Mobile Information Systems journal, Vol. 4, pp. 33-49, 2008.
- [9] Hu. Guoning, V. Deepak, "A Malware signature extraction and detection method applied to mobile networks", Performance, Computing, and Communications Conference, IPCCC 2007, IEEE Internationa, pp. 19 – 26, 2007.
- [10] Q. Yan, R. H. Deng, Yingjiu Li, and Tieyan Li, "On the potential of limitation-oriented malware detection and prevention techniques on mobile phones", International Journal of Security and Its Applications, Vol. 4, No. 1, pp.21-30, January, 2010.
- [11] D. Venugopal, Hu. Guoning, N. Roman, "Intelligent virus detection on mobile devices", Fourth International Conference on Privacy, Security and Trust, ACM PST, pp. 1-4, 2006.
- [12] C. Willems, T. Holz, F. Freiling, "Toward automated dynamic malware analysis using cwsandbox", IEEE Security & Privacy, pp. 32-39, March 2007.
- [13] L. Xie, X. Zhang, J. Seifert, S. Zhu, "PBMDs: A behavior-based malware detection system for cellphone Devices", 3rd ACM conference on wireless network security,WiSec10, ACM, pp. 37-48, March, 2010.
- [14] X. Zhang, Aciimez, O. Latifi, A. Seifert, S. Jose, "A rusted mobile phone prototype", Consumer Communications and Networking Conference, CCNC 2008. 5th IEEE, pp. 1208—1209, 2008.
- [15] X. Zhang, L. Xie, A. Chaugule, T. Jaeger, S. Zhu, "Designing system-level defenses against cellphone malware", 28th IEEE International Symposium on Reliable Distributed Systems, pp.83-90, 2009.
- [16] Liu L., Yan, G., Zhang, X., Chen, S., "VirusMeter: preventing your cellphone from spies", Proceedings of the 12th International Symposium on Recent Advances in Intrusion Detection, pp. 244-264, 2009.
- [17] G. Tuvell, D. Venugopal, Hu. Guoning, "Malware modeling detection system and method for mobile platforms", freepatentsonline.com, 2007.
- [18] V. Karyotis, A. Kakalis, and S. Papavassiliou, "Malware-propagative mobile ad hoc networks: asymptotic behavior analysis", Journal Of Computer Science And Technology, vol 23, pp.389-399, May 2008.
- [19] A. Schmidt, F. Peters, F. Lamour, "Monitoring smartphones for anomaly detection", Mobile Networks and Applications Volume 14, Mobilware08, ACM, Number 1, pp. 92-106, 2008.
- [20] M. Becher, F. C. Freiling, "Towards Dynamic Malware Analysis to Increase Mobile Device Security", SICHERHEIT, 2008
- [21] Mobile Malware Evolution: An Overview, Part 3. <http://www.securelist.com/en/analysis?pubid=204792080>

Identifying Nursing Computer Training Requirements using Web-based Assessment

Naser Ghazi, Gitesh Raikundalia
School of Engineering and Science
Victoria University, PO Box 14428
Melbourne, Australia

Janette Gogler, Leslie Bell
Austin Health
PO Box 5444, Heidelberg West
Melbourne, Australia

Abstract—Our work addresses issues of inefficiency and ineffectiveness in the training of nurses in computer literacy by developing an adaptive questionnaire system. This system works to identify the most effective training modules by evaluating applicants for pre-training and post-training. Our system, Systems Knowledge Assessment Tool (SKAT), aims to increase training proficiency, decrease training time and reduce costs associated with training by identifying areas of training required, and those which are not required for training, targeted to each individual. Based on the project's requirements, a number of HTML documents were designed to be used as templates in the implementation stage. During this stage, the milestone principle was used, in which a series of coding and testing was performed to generate an error-free product.

The decision-making process and its components, as well as knowing the priority of each attribute in the application is responsible for determining the required training for each applicant. Thus, the decision-making process is an essential aspect of system design and greatly affects the training results of the applicant. The SKAT system has been evaluated to ensure that the system meets the project's requirements. The evaluation stage was an important part of the project and required a number of nurses with different roles to evaluate the system. Based on their feedback, changes were made.

Keywords—component; Training Needs Analysis (TNA); Nursing Computer Literacy; Web-based Questionnaire.

I. INTRODUCTION

The health industry has come to rely heavily on the use of technology across all practice areas. The use of IT is an integral part of the industry's daily practices. Nurses must be well skilled in the various forms of IT, due to the wide-ranging duties that are inherent in their role. Austin Health, one of the leading health-care facilities in Melbourne, Victoria, utilises a variety of electronic systems at its three campuses [1]; i.e., a computer system at front desk to enter in-patient data, administrative duties on each ward, and systems for rostering, work shifts, and electronic incidents reporting. Computer literacy is a barrier for the effectiveness of completing required tasks as not all nurses possess a high level of computer literacy skill. Nurses Training is a key tool in achieving the objective of a skilled workforce. It therefore becomes essential for Austin Health to have adequate training measures in place, to ensure that nurses are in a position to understand and possess the requisite skills to work with the relevant electronic systems.

Austin Health recognises that each individual begins employment with an existing skill set that varies, depending on their education, level of experience and the scope of their

employment. Austin Health currently has a training method in place which sees all nurses, irrespective of their skill level and experience, undergo the same education. A system must therefore be developed to identify the current skill level of all current and incoming nurses in order to assess the appropriate type of training required to reach optimum knowledge and skill levels in electronic systems.

The System Knowledge Assessment Tool (SKAT) has been developed to meet Austin Health's objectives. The system is based on the use of an adaptive questionnaire, designed and developed in accordance with the principles of a Training Needs Analysis (TNA). The TNA involves a process of:

1. Determining the requirements of a particular nursing role
2. Establishing an individual's current skill and knowledge base
3. Identifying skill gaps and subsequent training solutions
4. Evaluating performance after training

A. Aims

The aim of this project is to replace the existing training method used by Austin Health with a new system that is efficient and cost effective for both the applicant and Austin Health. SKAT has been developed with regard to the current skill level and specific skill-set required for each nursing role.

The project aims to provide a range of benefits to Austin Health. SKAT provides:

- Efficient personalized training schedule and targeted training: SKAT provides a recommended training schedule which details the training required to meet the specific skill-set relevant to the applicant's role. Targeted training is considered most effective as it seeks to tailor the training to the needs of the applicant.
- Reduced training costs and training times: Training applicants only in areas where skill and knowledge gaps exist reduces training costs, as trainers will have less material to teach applicants. Training time will also be reduced as time will not be misused by teaching areas which applicants are already competent.
- Centralised skill progression monitoring: SKAT works as an effective skill progression tracking

tool with the capability of monitoring each applicant from the first stage of the TNA, through to identifying the applicant's current skill-set and knowledge base, and evaluating performance.

- An Information gathering tool: The survey contains a range of questions, including some which are intended for administration and information gathering purposes.

II. LITERATURE REVIEW

The nursing profession has always required the processing of information [2]. By the late 1960's the first computer systems were installed in hospitals. Initially, these computer systems were installed to complete orders, charges and patient billing, which were previously performed manually [3]. However, very quickly the power of computers became known, and they gained more presence on nursing wards. Computers in hospitals offer nursing a unique challenge to define and validate its own practice [4]. Since the 1970's, nurses have contributed to the purchase, design, and implementation of IT solutions, as it has become a major element of the nursing role [5]. Nursing data processed by the computer is referred to as *nursing informatics*. This term was created by Scholes and Barber, and refers to computer technology as well as the computer system used to process input data into information [6]. Today, NI has evolved to be an integral part of health care delivery and a differentiating factor in the selection, implementation, and evaluation of health IT that supports safe, high-quality, patient-centric care [3]. Nowadays hospitals in the western world depend on computer technology to manage day to day tasks. For this reason, nurses are required to be skilled in some specific areas of computer use to enable them to perform all the tasks that are required in the health-care industry.

A. Computer Literacy and Anxiety

Computer literacy describes the level of expertise and familiarity an individual has with computers. The term generally refers to the ability to use applications rather than a program.

In today's highly technological world, the use of computers either directly or indirectly may not always be an easy or enjoyable task. Various factors have been shown to cause levels of computer anxiety among users [7]. Beckers and Schmidt stated that "an experienced computer user understands enough about computers in order to use them, more or less independent of specific software packages, reasons for use and computer hardware features (p.786)." Thus computer experience can be seen as the sum of all computer-related events. These events include:

1. The amount of time spent using computers at home, office, school, work or any other place.
2. Using computer hardware such as a personal computer, printers, and scanners; and the use of software applications such as word processing programs, databases, programming, e-mailing, and downloading software from the internet.
3. The frequency of use of computers, e.g. once a day, regularly, etc. The more often computers are used; the

more responsive an individual will be to gaining skills and experience.

In an additional article in 2001 [8], Beckers and Schmidt proposed a model of computer anxiety. This model discusses six factors affecting an individual's levels of anxiety towards computers:

1. Computer literacy: An individual may feel nervous and uneasy when required to use a computer with no assistance. In contrast, a user with a high level of skill may feel much less anxious, as they have a higher level of experience, and thus more capable of completing tasks without assistance.
2. Self-efficacy: Self-efficacy is described as the confidence in users' capabilities to learn to use computers. Individuals who believe that they are incapable of using or learning how to use a computer will generally have higher levels of anxiety around computers.
3. Physical arousal in the presence of computers: Individuals who are not comfortable using computers, or who do not have many computer skills may show symptoms of anxiety in the form of sweaty palms, shortness of breath and seeming restless or unsettled.
4. Affective feelings towards computers: This factor identifies affective feelings such as likes and dislikes towards computers, by users. Generally, an individual who is not very skilled in the use of computers will have a negative feeling or dislike towards computers.
5. Positive beliefs: Identifies the benefits to society by using computers. In modern society, computers are perceived to be beneficial by saving time and effort when completing tasks.
6. Negative beliefs: Identifies negative factors which may increase user anxiety. Such effects of negative beliefs include feeling uncomfortable in the presence of computers, and may be related to previous unpleasant experience with computers, and thus invoke a negative connotation.

B. Computer Literacy of Nurses

The healthcare industry is constantly advancing, and care providers must recognise the need to sustain the required knowledge, skills, and resources required to communicate and manage information effectively in an electronic environment [9].

The computerised Health Information System (HIS) is expected to have great impact on health care practice in the years to come [10]. According to Dick and Steen, the HIS is "an essential part of technology for health-care today and in the future," and will lead to higher quality health-care, an increase in the scientific base of medicine and nursing, and a reduction in health-care costs [11].

Informatics competencies are an essential building block for evidence-based nursing practice [12]. Furthermore, the use of IT within the health-care industry is also expected to

improve patient care in a cost effective manner by saving time, improving accessibility of information and less paperwork [13].

Although nurses do not need a high degree of computer expertise, effective skills will save time and reduce human error, for example, nurses who operate computers proficiently can quickly access health-care related information using computers. Nurses may also be able to provide more appropriate and efficient care to their patients [14].

The American Nurses Association (ANA) proposed that NI's have the following skills: computer literacy skills, information literacy skills, and overall informatics competency. These skills are important to help prepare nurses to begin their practice, and are also important for experienced nurses and informatics nurse specialists [15]. Having a personal interest in IT is considered the best solution to computer anxiety, as it maintains an individual's confidence when using computers.

Personal interest in IT has become a topic of major interest [16]. Since personal interest in IT may have a positive impact on an individual's computer literacy, Hsu, Hou, Chang and Yen [14] have constructed a generalised research framework, resulting with conceptual framework of factors influencing computer literacy. Four hypotheses were proposed and are as follows:

1. Demographic variables have an influence on nurses' computer literacy.
 - The age of a nurse. Age has a significant influence on an employee's existing skill set and in some instances an inverse correlation to their training proficiency. For example, older nurses may not have had as much access to computer technology as their younger, more technology savvy colleagues. For this reason, it is understandable that younger nurses will generally have more computer skills, or even the ability to learn faster than older nurses.
 - The length of a nurse's work experience. Skills which an employee has gained from previous experience will play a role in their computing skills. However, this depends on the duration of their past experience and their position during previous employment.
2. The education level of a nurse. Computer education has a positive impact on nurses' computer literacy. An employee's education affects their skill set as better educated employees generally possess a more comprehensive skill set. An employee's level of education may correlate to their training proficiency.
3. Computer experience. Previous experience in the use of IT provides practice for the individual, which in turn creates a more competent and confident individual. Previous experience includes IT applications in the health-care industry such as e-filing, patient database management and information seeking through the internet.

4. Personal interest in IT. The nurses' personal interest in IT affects the skill set as it positively correlates with nurses' training proficiency. For example, the more interest in IT shown by a nurse, the more attention the individual will pay towards learning new skills; thus a higher level of information stored.

Nurses are engaged in two broad types of information seeking. First, information is sought to help make decisions about the care of individual patients. Secondly, information is sought about broader topics within nursing.

The Delphi study by Staggers, Gassert, and Curran [17] is the most relevant work to the authors' project, and has studied the informatics competencies of nurses in a detailed manner. The Delphi study concluded that the informatics competencies of nurses are expressed in four levels of practice, with different requirements for each level:

1. Beginning Nurse: A beginning nurse is required to possess important information management skills which are utilised in this role through working with large amounts of information and files which must be well maintained to ensure efficient management. Excellent computer technology skills are essential in this role as beginning nurses must use computer applications to complete day-to-day tasks and perform managerial duties in their relevant role.
2. Experienced Nurse: An experienced nurse must be well-skilled and trained in a specific area of interest such as public health, education or administration. Much like a beginning nurse; an experienced nurse must be highly skilled in using information management and computer technology skills to support the specific field of practice.
3. Informatics Nurse Specialist: An informatics nurse specialist is a registered nurse with a superior level of knowledge and skills specific to information management and IT. This role focuses on the education, administration, research and clinical practice information needs for the nursing role.
4. Informatics Innovator: An informatics innovator must be capable of performing research and generate theory, based on informatics. The informatics innovator must possess the foresight into possibilities, used to facilitate the development of informatics practice and research, leading to more effective time management to perform tasks. Thus, this level of nursing relies heavily on a high level of understanding and skills in IT and information management, and the understanding of the correlation of systems and disciplines.

The Delphi study identified the nursing informatics in four levels of practice. The results cannot be compared to those of the authors' project as the Delphi study identified nursing informatics in a more generalised manner, whereas the authors' project identifies the level of computer literacy of individual nurses, to generate a targeted training schedule.

C. Nurses Computer Education

Previous sections discussed assessing registered or graduate nurses' computer abilities, and existing skills. The following section discusses computer education and assessing undergraduate nurses' computer skills.

Information literacy development results from students' learning experiences within a number of subjects through their undergraduate years [18]. This ensures incremental development and enables students to transfer skills and knowledge across the undergraduate program at the same time. Wallace, Shorten, Crookes, McGurk, and Brewer proposed a number of strategies for introducing information literacy into the curriculum, for instance;

- Providing feedback regarding assessment information literacy skills and knowledge
- Articulating these skills in classrooms to students
- Developing subject aims and objectives
- Specifying the development of skills and knowledge involved
- Focusing on the process in addition to learning and providing the chance for discussing the process

When a teaching unit such as a department, school or faculty adopt them as a whole, these strategies are extremely effective. However, the adoption of such strategies across a teaching unit requires many academics to change the way they currently teach, therefore this approach occurs infrequently.

One of the critical issues facing the nursing profession is the lack of knowledge in IT and use of evidence-based practice, resulting from a serious lack of NI content in education and training programs. There is a pressing need for a computer driving license, which is a computer literacy certification program provided by the European Computer Driving License. The certification is recognised and supported by national governments and computer societies to assist nurses in becoming competent to use electronic information systems [19].

The policy of Nursing Informatics Australia [20], maintains that nursing education providers should ensure that nursing informatics education is included within the teaching materials in nursing programs. This will help establish the effective use of IT solutions and to support consumer-based care among new nursing graduates.

Nursing education providers should also provide students with opportunities to learn about the use of informatics tools to promote effective clinical decision-making as well as safety and quality. Providing students with these opportunities will enable the maximisation of workforce capabilities in the health-care field.

It is important that nursing educators provide an online learning community which can be accessed anywhere at any time, and provides a central resource for users. This increase in exposure to IT and systems within nursing curricula is important as it provides practice with various IT applications.

An advanced level of education in nursing informatics should be provided at a postgraduate level, to enable the high proficiency in areas such as informatics solution planning, development and management to provide an opportunity to develop more skills and knowledge in nursing informatics.

Nurses should be competent in nursing informatics as a main knowledge component before obtaining national registration and accreditation it is essential that nurses are competent in nursing informatics as the health-care industry is technologically advancing, and many aspects of the nursing role require a high level of computer skills. Nursing informatics education should be standardized with guidelines to measure the understanding of informatics competencies. This can be done by establishing national competency standards. Nurses must be educated in nursing informatics, as it is imperative that they keep abreast of informatics.

The nursing facility should be given enough support, to prepare for and enable the most effective teaching of health informatics. Health informatics is an integral part of the nursing role, and thus it is important that the nursing faculty is well prepared to teach health informatics. Special grants should be provided by governments to foster a supportive and effective educational environment in which nursing informatics can be taught.

Research topics should include the essential health informatics skills that a nurse should possess, the teaching methods and delivery of teaching to nursing students, as well as the most effective an efficient program design for nursing informatics subjects. Furthermore, the research topics should include the method of evaluation of the impact of the informatics education on clinical use.

In order to provide safe and effective nursing care, nurses must be capable of accessing an expanding information base, developing skills needed to manage the technology required for information retrieval and using this information appropriately to solve clinical problems. Ideally, the skills necessary to be a discerning user of information should be developed in pre-registration nursing programs.

III. SKAT REQUIREMENTS

This project is designed and developed for use by Austin Health. There are approximately 750 public hospitals and 500 private hospitals in Australia, accommodating more than 50,000 hospital beds for patients, over 7000 staff, of which there are 3200 nurses. Austin Health is a major health-care institution in Victoria and comprises three facilities: Austin Hospital, Heidelberg Repatriation Hospital and the Royal Talbot Rehabilitation Centre [1].

A. Austin specific problem

With an increasingly large amount of new staff employed at Austin, it is understandably difficult to track the level of computer skills of each staff member. This creates a problem when new staff members undertake generic training that may not be required, as the individual may already have the knowledge and skills and only need to understand the manner in which the processes are used at Austin.

Consequently, with minimal training effort they can become proficient users of the system. However, the challenge lies where there are other staff members with little or no skills and knowledge, and therefore need a greater level of in-depth training. Austin has identified a benefit in that targeted training has the potential to reduce costs and lead to improved outcomes in a range of areas.

B. Project Goal

The goal of this project is to provide a computer-based test that assesses the computer literacy level of each nurse, thereby determining the learning modules required to meet the set competency levels. The testing application is adaptive; future questions will be determined by previous answers provided by the nurse.

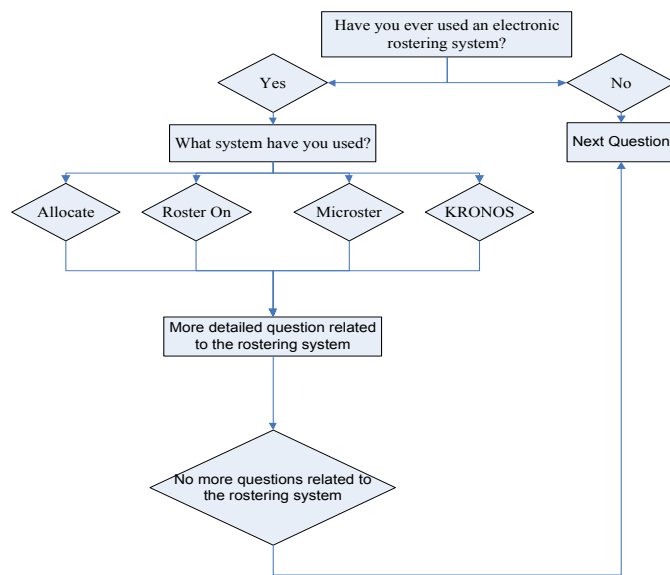


Figure 1: KRONOS Example

Fig. 1 demonstrates an example of the adaptive testing application. In this example, the main question 'Have you ever used an electronic rostering system?' must be answered by the applicant in order to progress to the next question.

If the applicant's answer is 'No,' the test questions will change to suit an applicant with no previous rostering experience. However, if the applicant's answer is 'Yes,' the application will display more detailed sub-questions to effectively probe the applicants past history with computer systems, and therefore obtain as much information as possible from minimal questions. The course of questioning will test for general computer skills, skills within specific programs and software packages and other systems used to determine the level of the applicant's technical knowledge.

The application will process each answer and provide a collective result indicating the technical skill level of the applicant and the training modules required to meet the required skill set. The results for each assessment will be saved in the employee's respective file within the hospital database.

C. Austin's Nursing Roles

Austin Health Institute employs eight different nursing positions.

1) *Nurse Unit Manager (NUM)*: A NUM is a Registered Nurse (RN) who supervises nurses in a specific unit of the hospital. This role involves managing the care of patients and the responsibilities of nurses, budget management and rostering of staff. The role of a NUM also involves the management of the Patient Services Assistant and the ward clerk [21].

2) *Assistant Director of Nursing (ADON)*: The ADON is a senior management position within Austin Health. Most of the ADONs are the after-hours site managers responsible for the running of the hospital in the evening, overnight and on weekends. This role requires exceptional communication skills to manage bed occupancy, staffing issues, patient and relative issues that arise as well as emergency and disaster management. The ADON also has a role to report back to the Deputy Director of nursing [22].

3) *Associate Nurse Unit Manager (ANUM)*: The ANUM has significant responsibilities. In the absence of the Nurse Manager, ANUM has the responsibility of undertaking tasks delegated by the Nurse Manager. As members of the hospital's middle management team, the ANUM is usually the nurse coordinator of the shift and takes on a portfolio of responsibility in the ward or the department [23].

4) *Clinical Nurse Specialist (CNS)*: A CNS has advanced practice skills and holds graduate degrees (master's or doctorate) in a specific field of medicine such as neonatal, developmental disabilities, diabetes, oncology or pediatrics [24].

5) *Clinical Nurse Consultant (CNC)*: A CNC provides direct patient care and creates and implements health care protocols. The CNC ensures that clinical employees abide by government laws and regulations when performing tasks [25].

6) *Clinical Support Nurse/ Continuing Nurse Educator (CSN/CNE)*: A CSN provides teaching and support to Graduate Nurses (GN), year program participants and other staff in a manner consistent with achieving learning outcomes set by the Clinical Nursing Education Department (CNED). This role is ward based and supports graduate year nurses in their clinical roles.

A CNE is a Registered Nurse who coordinates and ensures appropriate delivery of specific education program components, short courses or study days; to achieve outcomes set by the CNED [26].

7) *Registered Nurse (RN)*: An RN is a qualified nurse with a degree in nursing. The role of an RN is much broader than the other seven nursing positions, and mainly involves first-hand interaction with patients, and no managerial role [27].

8) *Enrolled Nurse (EN)*: An EN is a second-level nurse who offers medical care under the supervision of an RN [28].

D. Austin's Existing Systems

Austin Health uses a number of systems which nurses have to be trained to use, in order to fulfill the responsibilities for the role they occupy.

1) *KRONOS*: Electronic rostering or e-rostering is the process of using the computer's power to the task of rostering. Essentially, e-rostering considers the requirements, skills, shift history, preferences of staff members and creates a schedule that not only meets the requirements but also complies with laid-down rules of scheduling.

KRONOS is the computerised roster and timecard management system used by Austin Health. It has the capability for staff to make electronic requests for leave, accruals can be viewed and reports can be printed [29].

2) *TrakHealth*: TrakHealth develops and markets TrakCare, Austin Health's patient administration system. The electronic Patient Records (ePR) in TrakCare are operational. It uses the database identifier or the Patient Master Index (PMI) for all patients receiving services at the hospital to create and store patients' unit medical record numbers. It also provides a full range of clinical, administrative, lab and community care capabilities, unified in a single data repository [30].

3) *Scanned Medical Record (SMR)*: SMR is a web-based system that is used within Austin Health for scanning the patients' printed medical records. The records are scanned and saved in the database in a Portable Document Format (PDF) format once the patient is discharged. The records can be monitored, printed or emailed from anywhere in the Hospital and at any time [31].

4) *Online Radiology System*: The Picture Archiving and Communications Systems (PACS), is the Online Radiology System used by Austin Health. PACS allows authorised users to access radiological images online to multiple users, 24 hours a day, on all Austin Health sites. PACS uses a radiology web browser called AGFA Web1000 to view the radiology images on the computer [32, 33].

5) *Risk Management*: RiskMan is a system which electronically records all clinical and non-clinical accidents or misses. Austin Health is committed to promptly reporting and responding to incidents and hazards, to ensure the best possible outcomes for all people affected and to identify and control future risk to staff and members of the community [34].

6) *Budgeting system*: PowerBudget is a financial software application for the NUM used to anticipate, forecast and manage staff costs and consumables in each ward or department. The system provides intelligent budget calculations using cost driver relationships generating instant results based on the user's requirements [26].

7) *Electronic medication administration system*: Cerner Millennium is the new electronic medication administration system that has been implemented within Austin Health. This system was initiated by the Victorian Department of Health as part of the HealthSMART program; a state-wide plan aiming to

modernise and replace IT systems throughout the Victorian Public Health-care Sector [26].

8) *Health-e Workforce Solutions (HWS)*: HWS is a software system designed to facilitate the management and analysis of the nursing workforce, especially the management of replacing staff on sick leave, roster vacancies and annual leave. The application determines the requirements for each area related to nursing classifications and roster requirements. This provides information on the Exchange Traded Funds (EFT) requirement for the respective ward or area and alerts managers to a shortfall in staff and level required.

Nurses in charge of the shift are able to enter staff sick leave and replacement requirements, while managers have the capability to access greater functionality to ensure that rosters are created with the correct level of staff for each shift [35].

9) *Comprehensive Human Resource Integrated Solution (chris21)*: Chris21 is the Human Resource and Payroll Management system used within Austin Health. Chris21 builds a complete history of every employee which creates the ability to build HR and payroll strategies based on certain requirements. The system delivers a dedicated, multi-functional and highly-flexible HR solution [36].

E. Overall System Description

SKAT gauges the level of computer skills of each individual and comprises two parts; Assessment component, and Administrative component. These two parts work together to assess the level of computer literacy, based on the input provided by the applicant.

1) *Assessment Component*: The Assessment component of the system allows applicants to enter their personal details and choose their appropriate role. SKAT uses this information to determine which corresponding questionnaire will be administered to the applicant. The questionnaire encompasses three main categories; General IT Skills, Professional IT skills, and Hospital Systems skills.

The General IT Skills category includes a number of sub-categories to obtain a thorough understanding of the individual's abilities in using various IT applications

- General information
- Social networking
- Search engines
- Personal use
- Internet browsing

The category of Professional IT Skills helps to identify which computer applications have been used previously by the individual; to gauge the level of competency in the use of IT at work. This category covers the following sub-categories:

- Emailing skills
- Microsoft Word
- File management
- Microsoft Excel
- Microsoft PowerPoint

The Hospital Systems category focuses more specifically on IT systems which are currently used in the health-care system, and assesses the level of exposure to these systems to gauge the individual's experience. This category explores the following sub-categories:

- Electronic rostering
- Electronic casual staff management system
- Electronic risk management
- Electronic budgeting
- Clinical systems
- Electronic medication administration
- Mental health
- Residential aged care
- Community care

The applicant must answer all questions before submitting the questionnaire; the software then generates a report with the recommended training for the particular role of the individual.

2) *Administrative Component:* The administrative component presents the System Administrator with the ability to control the functionality of the software. The system provides an interface tailored to each user type, according to input provided. Elements of the questionnaire are adaptive, that is, the next question presented will be determined by preceding answers.

The course of questioning tests for general computer skills, health business and clinical systems skills used by different levels of registered nurses within Austin Health, as part of the daily requirements, to determine the level of the applicant's technical knowledge. This will enable training to be specific to each applicant, and result in time efficient, and cost efficient training.

Fig. 2 summarizes the Austin Health Survey System Use Case. It illustrates the functions that each of the three user types can perform, and shows the differences between the three user roles. The Figure shows that the Applicant can only perform one task, whereas the System Administrator can administer all functions but one. As shown, Managers and System Administrators have shared access to certain functions.

IV. SKAT DESIGN

The design of SKAT system encompasses three designing stages; the Questionnaire design, Database design, User Interface (UI) design, and Decision Making design.

A. Questionnaire Design

The questionnaire encompasses three main categories, with 19 sub-categories. The multi-level design of the questionnaire simplifies the display of the questionnaire, by grouping questions into sub-categories, and these sub-categories into more general categories. This hierarchical design is a much simpler way of forming the questionnaire, as questions are grouped in relevant categories and sub-categories to allow more efficient navigation through the questionnaire.

B. Database Design

The questionnaire database controls the required information for the applicant from the start to completion of the questionnaire, and manages the following steps:

- Personal details of the applicant
- The role of the applicant
- The most appropriate survey required
- Survey questions
- Corresponding possible answers
- Save the report for each survey

The Database Entity Relationship (ER) diagram, Fig. 3 illustrates an important part of designing the database. It depicts the main entities required for the database with the corresponding attributes for each entity. The figure also illustrates the relationship between the entities, as an example, one or more (1..*) applicant might have the same position (1..1).

C. User Interface Design

HTML documents were developed as mockups of the system's user interfaces which are designed as a model of the software to show the appearance of the application's design. The development of the UI design reached its final stage after a number of changes were made, including changes to the UI, grammar, categorical structure, and questions, based on feedback from the project supervisors at Austin Health. The UI Mockups are used as templates to implement the application.

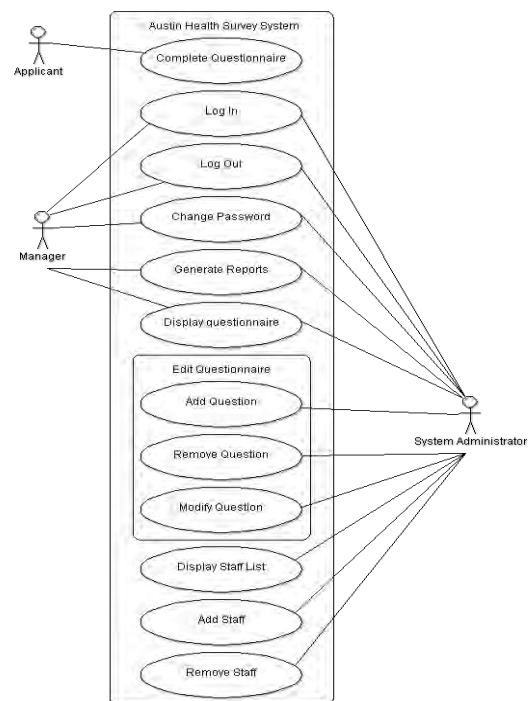


Figure 2: Austin Health Survey System Use Case

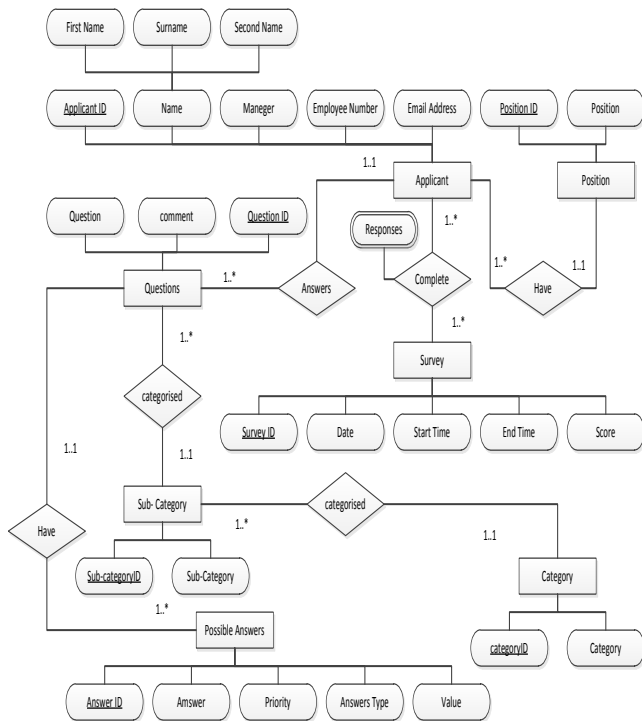


Figure 3: Entity Relationship Diagram

D. Decision Making

The Decision Making component is the process responsible for determining the required training for each applicant. The system offers three versions of the questionnaire, based on the applicant's position. Training recommendations vary, according to the needs identified through the applicant's responses and their role.

A training priority procedure will apply, and will have two approaches:

- Setting a priority level for a specific field or possible answer. This approach is useful, for example where an applicant selects 'no' in a question regarding the applicant's use of a particular system. As a result of no previous experience in this system, SKAT identifies that a higher priority of training is necessary.
- Answers will be rated from 1 to 5, and will be determined according to the applicant's score in the questionnaire. A lower score indicates lower skills, making training requirements a higher priority for that applicant. However, a higher number indicates a higher level of skills, making training requirements a lower priority for the applicant.

Each role will have a Decision Support System (DSS) for priority training as shown in Table 1.

TABLE 1. DECISION MAKING PRIORITY LIST

Priority 0 (P0)	no training
Priority 1 (P1)	training must be within 2 weeks from employment
Priority 2 (P2)	training must be within 4 weeks from employment
Priority 3 (P3)	training must be within 6 weeks from employment

The DSS shown in Table 1 illustrates the four priority levels determined through the results of the questionnaire. The priority time frames have been set by Austin Health, as they are the most appropriate intervals for the roles and positions available.

V. SKAT SYSTEM

The SKAT system consists of 24 web pages including the Master page, Header page, Questionnaire starting page, Applicant's information page form, Questionnaire pages, Closing page, and Report Page. SKAT is a web-based application which will run on Austin Health's local server (Microsoft Windows Server 2003). SKAT will be accessed and used by multiple users simultaneously, to facilitate various functions. The programming languages used to program the SKAT were ASP.net, HTML, and C Sharp (C#).

The Author used Microsoft Visual Studio 2010 Professional, Framework 4.0. ASP.net was chosen as the server at Austin Health conflicts with Java and Java Script. In view of the fact that Austin Health uses Microsoft products, this was the main reason for choosing Microsoft Visual Studio 2010 to develop the software. Developing the software in ASP.net will increase the systems compatibilities, and reduce conflicts between programming language and servers.

The first author used the milestones principle, in which the coding is tested after each major step or after adding a new function. This method helps to correct all developing errors in each stage, which will increase the chance of producing bug-free software.

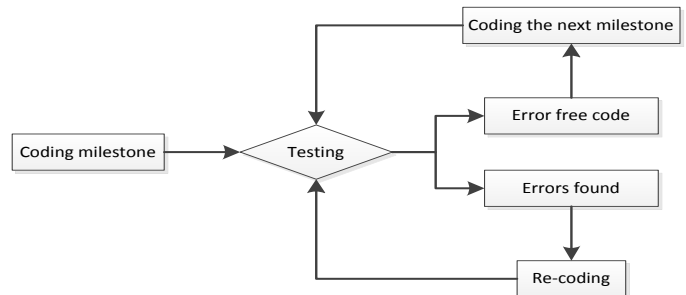


Figure 4: Testing milestone

The process of testing as shown in Fig. 4 begins by testing at each milestone stage, and if the code contains no errors, the next milestone will be tested. However, if errors are found in the code, the author must re-code and test, to ensure that no further errors exist in the code.

A. SKAT Pages

The primary page a user will view upon beginning the survey is the SKAT Welcome page. This page serves the purpose of providing a short introduction to SKAT, including a brief summary of the components of the questionnaire, and question type. It also provides simple instructions about starting the questionnaire and completing it.

The Applicant information page shown in Fig. 5, requires the user to enter general personal details such as position, name, manager, employee number and email address, and serves the purpose of identifying the user.

APPLICANT INFORMATION
Please fill in your personal information

Position: Please select the classification that best describes your role

Manager: Please select your Manager

First Name: Surname: Second Name:

Employee Number: Email Address:

First Name, Surname, Position, and Email Address. Are required

Figure 7: Applicant's Information

The information entered will be gathered to ensure that the appropriate surveys will be available to the user, and to record information about start and finish times of questionnaires, and applicants' details.

GENERAL IT SKILLS

GENERAL INFORMATION

How often do you use the computer ?

Frequently Regularly Occasionally

Where do you use computers ?
Please tick those applicable

Home Work University Library

You use computers for ?
Please tick those applicable

Work Study Entertainment

Figure 5: General IT Skill's page

Fig. 6 illustrates the first set of questions that applicants must answer. These questions are organised into sub-categories. The questions have been designed to obtain general information about the level of computer use quickly. The applicant must answer all questions before pressing on the 'Next Page' button to move to next page.

PROFESSIONAL IT SKILLS

FILE MANAGEMENT

Saving documents
Please rate a response that best describes your current level of skill or ability.

Excellent Good Average Fair None

Saving documents to folders
Please rate a response that best describes your current level of skill or ability.

Excellent Good Average Fair None

Organizing and naming files
Please rate a response that best describes your current level of skill or ability.

Excellent Good Average Fair None

Searching for a document
Please rate a response that best describes your current level of skill or ability.

Excellent Good Average Fair None

Figure 6: Professional IT Skills page

The Professional IT Skills page, shown above in Fig. 7 presents a set of questions from a different category. The responses given to these questions determine the applicant's ability to complete basic computer tasks, such as saving, maintaining and locating files. The questions on this page have been set up to provide the most amount of information as efficiently as possible. Upon completion of the questions on this page, the applicant must press the 'Next Page' button to continue to the following sub-category.

HOSPITAL SYSTEMS

ELECTRONIC CASUAL STAFF MANAGEMENT

Have you used an electronic casual staff management system?

Yes No

Which of the following systems have you used?
Please tick those applicable

HWS CASCOM Allocate KRONOS Other

You have used the system as?
Please tick those applicable

Employee Manager Super-user

Figure 8: Hospital Systems page

The last set of questions is displayed in the Hospital Systems page as shown in Fig. 8. These questions are more specific to the role of the applicant, as they seek information about computer systems and the area of work the systems have been used in. As this is the final page of this category, the applicant must press the Submit button to complete this section.

Upon completion of the questionnaire, a closing page is displayed, with a message to thank the applicant for completing the questionnaire. This page also contains a link to the corresponding report for the applicant. The report provides details of answers provided for the individual questionnaire.

System Knowledge Assessment Tool SKAT		Applicant's Report	
Applicant's Name:	Naser Ghazi	Position:	NUM
Start Time:	3:52 PM	Email Address:	naser.ghazi@live.vu.edu.au
End Time:	4:11 PM	Date:	10/26/2011
		Employee Number:	3804093
Category	Sub-Category	Question	Applicant's Response
General IT Skills	General Information	How often do you use the computer ?	Frequently
		You use computers for ?	Entertainment
	Internet browsing	Adding a site into favorites	Fair

Figure 9: Applicant's responses Report

The Applicant's Report provides a list of answers provided by the applicant, as well as an overview of the applicant's personal information such as name, email address, position and employee number, displayed in the information bar. The report also details the date, start and finish times of the questionnaire.

The page contains a toolbar with options to skip pages, refresh the page, zoom in or out, or to find keywords by using the search function. The toolbar also provides an option to save the report as a PDF, Microsoft Word document, or a Microsoft Excel document, as well as printing the report.

B. SKAT Validation

Two types of validation were applied to SKAT. The first type of validation is to check SKAT’s navigation, to ensure that all mandatory questions are answered before navigating to the next page. The second type of validation involves checking user inputs to ensure that they meet the requirements of the field, for example a particular answer field may only allow numerical digits to be entered. Validating this field ensures that the applicant cannot enter text into the field.

Figure 10: First Name field Validation

A first name field validation was performed on the Applicant Information page to validate the applicant’s input. In Fig. 10, the applicant entered numeric figures in the ‘first name’ field. Upon validation of this page, an error message appears, notifying the applicant that they must enter information into the required field. This field does not allow numerals, punctuation marks and symbols to be entered. Only upper case or lower case letters of the alphabet may be entered.

Required answer validation was performed on the questionnaire pages to ensure that the questionnaire will not proceed unless all questions have been answered. If a question is not answered, an error message will be displayed above the relevant question, advising the applicant that an answer is required.

Figure 11: Required Answer Validation

Fig. 11 shows an example where the last question was not answered, and an error message prompting the applicant to answer the question.

VI. SKAT EVALUATION

Evaluation is an important element of system development. It ensures that the system developed meets the client’s requirements. Furthermore, with regard to SKAT, it provides information about the effectiveness and efficiency of the system, based on the experiences by nurses testing the system. Evaluation is also a critical part of development, as it enables nurses testing the system to provide suggestions and comments to improve SKAT.

The most appropriate method of evaluation was identified as a simple questionnaire, due to time and staff constraints present within Austin Health. This questionnaire is the most efficient way of obtaining feedback about SKAT, and the most suitable method for Austin Health, as it also enables testers to leave comments to improve the system, in addition to verbal suggestions during the completion of SKAT.

The evaluation questionnaire, as administered to test subjects is as follows:

Figure 12: Evaluation Questionnaire

Fig. 12 illustrates questionnaire is a suitable method of evaluating the system in terms of ease of use from the applicant’s perspective. It also provides an opportunity to identify improvements which can be made to SKAT.

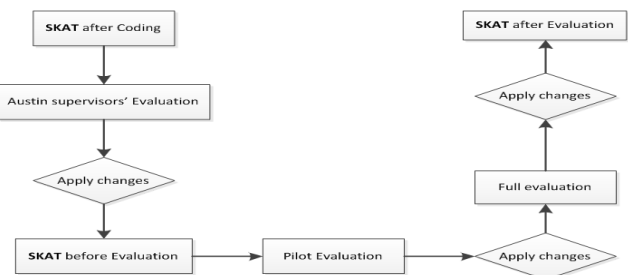


Figure 13: Evaluation Process

Fig. 13 illustrates the flow of steps taken in the process to evaluate SKAT. The first step in this process is the Austin supervisors’ Evaluation, the second step is the pilot evaluation, and based on these results, changes were applied to SKAT. The

third and final step is a full evaluation, which involved applying final changes to SKAT, based on feedback provided by evaluators, before delivering the final version of SKAT to the client.

A. Pilot Evaluation

Upon completion of programming and testing, a User Acceptance Test (UAT) was performed on SKAT. When evaluating the system, the Austin supervisors completed the SKAT as applicants and identified grammatical errors which needed to be corrected. Grammatical errors are the only changes which can be made at this stage, before the pilot evaluation is conducted. The errors were discussed with the author and changes which needed to be made were identified.

A pilot evaluation was then conducted, using nurses in the NUM and CNS roles. The NUM participated in the pilot evaluation, as the NUM is presented with the most questions in the questionnaire, as it is a more senior role. The pilot evaluation concluded that more grammatical errors were identified, and some questions needed to be re-worded to improve understanding by applicants. The NUM and CNS provided suggestions for changes to the SKAT, and as these suggestions were also raised by the Austin supervisors after the UAT, the changes were applied.

After the CNS and NUM undertook the pilot evaluation, repetitive and unnecessary questions were identified, and consequently discussed with Austin supervisors who then took necessary steps to amend the questionnaire.

Table 2 explains reasons for the removal of thirteen questions in SKAT. It summarizes information about the sub-categories, questions and reasons they were removed from the questionnaire.

TABLE 1. REMOVED QUESTIONS

Question	Sub-Category	Reason of removing
Do you use social networking applications?	Social Networking	Repetitive
Face book	Social Networking	Repetitive
Twitter	Social Networking	Repetitive
MSN-Messenger	Social Networking	Repetitive
Yahoo-Messenger	Social Networking	Repetitive
Skype	Social Networking	Repetitive
Do you use search engines?	Search Engines	Repetitive
Google	Search Engines	Repetitive
Bing	Search Engines	Repetitive
Yahoo	Search Engines	Repetitive
MSN	Search Engines	Repetitive
Pay bills online?	Personal Use	Repetitive
Opening a file into the browser?	Internet Browsing	Irrelevant
Text Editing?	Microsoft Word	Repetitive
Saving documents to folders?	File Management	Repetitive

The questions in the two sub-categories; Social Networking and Search Engines, as well as others were removed from SKAT as they do not affect the applicant's overall score in the survey. These two sub-categories provide a resource for information gathering purposes within Austin health.

TABLE 2. ADDED QUESTIONS

Question	Sub-Category
Social Networking applications, E.g. (Face book, Skype, or Twitter)	General Information
Search Engines , E.g. (Google, Bing)	General Information

Table 3 shows that two questions were added to SKAT, under the General Information sub-category. These have been moved from two other sub-categories, as discussed under Table 2. These two sub-categories were amalgamated into only two questions in the General Information sub-category as the original questions were deemed too repetitive. The new categorizing of these two questions provides a much simpler way of gathering this information.

Figure 14: General Information Sub-Category after pilot evaluation

Fig. 14 depicts the General Information page after the UAT and pilot evaluation were conducted. Austin supervisors, NUMs and CNSs identified that some questions in the Social Networking and Search Engines pages were repetitive and unnecessary as they were irrelevant to the industry. Consequently, these two sub-categories were incorporated into the General Information sub-category. Two sub-categories have been removed completely from the questionnaire, thus reducing the number of sub-categories in the SKAT from 19 to 17, as very detailed information about these sub-categories is not required. The total number of questions has also been reduced from 87 to 74 questions.

Figure 15: yes/no questions after pilot evaluation

Fig. 15 illustrates the changes made according to the UAT and pilot evaluation. The option 'None' was added to questions in all sub-categories starting with a yes/no question. Previously, the questionnaire did not list 'None' as a response option; hence the applicant was obliged to choose an answer before proceeding to the next sub category, even if the answer was not applicable.

B. Evaluation

The full evaluation stage is the final step in the evaluation process before delivering the complete system to the client. This step is considered a re-evaluation of the changes made to the system after the pilot evaluation to ensure that the system works well, and can be easily understood by the applicant.

The full evaluation process used in SKAT involved the participation of five nurses with the following roles: NUM2, NUM3, ANUM, CNS2 and RN. These nursing staff tested the system and completed the questionnaire sample, and generally had positive attitudes towards the system. Based on the full evaluation feedback, the following changes were made to SKAT.

ELECTRONIC CASUAL STAFF MANAGEMENT
Have you used an electronic casual staff management system?
If your answer was "No", Please select "None", for the following questions in this section.

Yes No

Which of the following systems have you used?
Please tick those applicable

HWS CASCOM Allocate KRONOS None

In which role have you used the system?
Please tick those applicable

Employee Manager Super-user None Change radio buttons to check boxes

Figure 16: Applicant's role after Evaluation

After the evaluation process, the Author and Austin Health supervisors agreed that it is simpler and easier for applicants to comprehend if they changed the possibility to have more than one answer option, or to tick all answers that apply. NUM2 found this to be an issue. Consequently, changes in coding were made to change radio buttons into check boxes to allow more than one answer to be selected. Fig. 16 demonstrates the changes to coding, to enable check boxes to allow more than one response to be submitted. This provides more detailed information about the applicant, as it allows for as much information as possible to be selected.

Changing fonts
Please rate your skill-level on the following scale.

Excellent Good Average Fair None

Inserting a picture or a chart into a document
Please rate your skill-level on the following scale.

Excellent Good Average Fair None

Changing page layout and orientation
Please rate your skill-level on the following scale.

Excellent Good Average Fair None

Your Answer is required!

Figure 17: Page Validation after Evaluation

An additional validation error message was added to alert the applicant of required answers. As a result, the page now displays one message at the top, and one message at the very bottom of the page. These two error messages advise the applicant that a mandatory question has not been answered, without having to scroll up through a long page to check messages. Instead, the applicant will see any error messages at the bottom of the page near the 'Next Page' button upon attempting to submit.

VII. CONCLUSION

The use of modern technology in the health industry demands that employees are given adequate training to ensure a productive work environment. Austin Health recognizes that effective training measures should be in place to ensure that nurses have the requisite skills to work with the relevant electronic systems.

Austin Health observed that the current method used for training nurses was time consuming, uneconomical and generally inefficient. In particular, the disparity between nursing roles meant that a one size fits all training approach was not appropriate, as it failed to take into account the different training needs of each nurse.

In response to these concerns, Austin Health considered replacing its existing training method with a new system that would identify the current skill level and specific skill-set required for existing and new nurses. Taking into account the various alternatives to the current method, Austin Health considered the development of a web-based survey as an appropriate training solution. Through ongoing consultation with the first author and further research gathered, SKAT was developed to meet Austin Health's objectives. The system uses a web-based adaptive questionnaire designed and developed in accordance with the principles of the Training Needs Analysis (TNA). The ultimate result is a training system that provides efficient targeted training, through personalized training schedules within time and costs constraints.

The design of SKAT developed through a process which began with the design of an efficient database system, defining a data type for each attribute and entering data into the database. This process follows through to the coding of SKAT, which is inextricably linked to the design of the system, as coding is applied according to the design.

Upon completion the coding stage, the third and the fourth authors provided an evaluation of SKAT, and changes were applied according to their feedback. A pilot evaluation was undertaken by staff in two different nursing roles, and their suggestions were discussed and applied to the system. This process concluded with a full evaluation of SKAT, and involved applying final changes to the system based on feedback provided by evaluators, before delivering the final version of SKAT to the client.

VIII. FUTURE WORK

Although SKAT was designed, implemented and evaluated after its completion, there still many functions which could be added to the system.

- Adding a progress bar to each page of the questionnaire. This function helps the applicant to track progress throughout the completion of the questionnaire.
- Applying SKAT to different professions. This system can be easily applied to other professions by implementing appropriate requirements to form a relevant questionnaire.
- Automatic sending of applicant's results to managers. This function will enable applicant's reports to be automatically attached to an email, and sent to the applicant's manager.
- SKAT can automatically connect to other systems. The system can share information about results generated in SKAT by connecting with other systems, such as Chris 21, the Human Resources system at Austin Health, which can book training sessions for applicants based on results generated from their questionnaire.
- SKAT will automatically send each applicant an email with a link to begin the survey. Individuals can then access and complete the survey from any location such as home or work, and all responses will be saved automatically on the server.

Although some of these features are not critical, they would all clearly provide a great amount of benefit to SKAT if implemented.

ACKNOWLEDGMENT

There are many individuals who contributed to the successful completion of this thesis through their support, advice or participation. We must single out Sarah Tawil for her support and assistance in checking and advising writing the thesis. We are also grateful to Russell Paulet for his assistance in programming and technical support.

REFERENCES

- [1] http://www.austin.org.au/about_us/
- [2] F. Nightingale, *Notes on Nursing: What It Is and What It Isn't*. Dover, New York, 1859.
- [3] J. Murphy, "The Intersection of Nursing, Computer, and Information Sciences. Nursing Economics," *Nursing Informatics*, 2010, vol. 28, p. 3.
- [4] J. Kelly, "Computer in hospitals nursing practice defined and validated," *Department of Nursing, New York University Medical Centre*, 1983, vol. 83, pp. 195-210.
- [5] M. Ball, K. Hannah, *Using computers in nursing*. Reston, VA: Reston Publishers, 1984.
- [6] M. Scholes, *the Impact of Computers in Nursing: an International Review*. North-Holland, Amsterdam, 1983.
- [7] J. J. Beckers, H. J. Schmidt, "Computer experience and computer anxiety," *Computers in Human Behavior*, 2003, vol. 19, pp. 785-797.
- [8] J. J. Beckers, H. J. Schmidt, "The structure of computer anxiety: a six-factor model," *Computers in Human Behaviour*, 2001, vol. 17, pp. 35-49.
- [9] Bureau of Labor Statistics, *Occupation outlook handbook 2004*, 5 edition Registered nurses, 2003.
- [10] E. Goorman, M. Berg, "Modelling nursing activities: electronic patient records and their discontents," *Nursing Inquiry*, 2000, vol. 7, pp. 3-9.
- [11] R. E. Dick, E. Steen, *The Computer based Patient Record: An Essential Technology for Healthcare*. National Academy Press, Washington D.C., 1991.
- [12] S. Bakken, "An informatics infrastructure is essential for evidenced-based practice," *Journal of the American Medical Informatics Association*, 2001, vol. 8, pp. 199-201.
- [13] R. Gururajan, S. Murugesan, J. Soar, "Introducing Mobile Technologies in Support of Healthcare," *Journal of Information Technology Management*. 2005, vol. 18, pp. 12-18.
- [14] W. W. Jiang, W. Chen, Y. C. Chen, "Important Computer Competencies for the Nursing Profession," *Nursing Research*, 2004, vol. 12.
- [15] American Nurses Association, "Scope and standards of nursing informatics practice," Washington, DC: American Nurses Publishing 2001.
- [16] H. M. Hsu, Y. H. Hou, I. C. Chang, D. C. Yen, "Factors Influencing Computer Literacy of Taiwan and South Korea Nurses," *J Med Syst*, 2009, vol. 33, pp. 133-139.
- [17] N. Staggers, C. A. Gassert, C. Curran, "Informatics Competencies for Nurses at Four Levels of Practice," *Nursing Education*, 2001, vol. 40.
- [18] M. C. Wallace, A. Shorten, P. A. Crookes, C. McGurk, C. Chris Brewer, "Integrating information literacies into an undergraduate nursing programme," *Nurse Education Today*, 1999, vol. 19, pp. 136-141.
- [19] "European Computer Driving License", 2011, from <http://www.ecdl.org/index.jsp?p=93&n=94>.
- [20] Nursing Informatics Australia, (NIA), 2001. Position Statement. Integration of Nursing Informatics into Nursing Education.
- [21] L. Friend, "Nursing Unit Manager Job Description", 2011, from http://www.ehow.com/about_6331258_nursing-unit-manager-job-description.html#ixzz1VCL67nOP
- [22] J. Robertson, "Assistant Director of Nursing Services Job Description", 2011, from http://www.ehow.com/about_6330371_assistant-nursing-servicesjobdescription.html#ixzz1VCOJUuv
- [23] W. Thibodeaux, "Assistant Nurse Manager Job Description", 2011, from http://www.ehow.com/about_6551492_assistant-nurse-manager-job-description.html#ixzz1VLDYgZ8m
- [24] L. Kelchner, "Job Description of a Clinical Nurse Specialist", 2011, from http://www.ehow.com/about_6739002_job-description-clinical-nurse-specialist.html#ixzz1Vou0Ok9S
- [25] M. Codjia, "Clinical Nurse Consultant Job Description", 2011, from http://www.ehow.com/facts_6803456_clinical-nurse-consultant-job-description.html#ixzz1VCRG5F9k
- [26] J. Gogler, Assistant Director of Nursing, Nursing Informatics, Austin Health, 2011.
- [27] "Job Description for an RN Nurse", 2011, from http://www.ehow.com/facts_4840477_job-description-rn-nurse.html#ixzz1VpMWwwnl
- [28] J. Price, "Job Description of an Enrolled & Endorsed Nurse", 2011, from http://www.ehow.com/facts_7207795_job-description-enrolled-endorsed-nurse.html#ixzz1WHwChC9c
- [29] Kronos, Australia Pty Limited is a Government Endorsed Supplier, 2011, form <http://www.kronos.com.au/about-kronos/about-kronos-australia.aspx>
- [30] TrakHealth Pty Ltd, 2004, from <http://www.consensus.com.au/SoftwareAwards/CSAarchive/CSA2006/TrakHealth.htm>
- [31] InfoMedix, wholly-owned subsidiary of Object Consulting Pty Ltd – Australian Software Development Company, 2011, from <http://www.infomedix.com.au/>
- [32] Accenture, Medical Imaging, 2011, from http://www.accenture.com/au-en/landing-Pages/health-public-service/Pages/healthps-au-find-the-optimal-medical-imaging-solution-with-healthanalytics.aspx?c=con_auglohpsgs_1210&n=g_medical_imaging/_a_0_k/pacs&KW_ID=33bde296-636a-51e9-42c7-000034b89524
- [33] MEDIVISION, Health-Care informatics, "IMPAX, Agfa-Web 1000", 2006, from <http://www.medivision.co.il/agfa-WEB1000.htm>
- [34] RiskMan International (RMI), 2011, from <http://www.riskman.net.au/>
- [35] Microsoft Partner, Health-e Workforce Solutions, 2011, from <http://www.healthewfs.com.au/>
- [36] Frontier Software, Comprehensive Human Resource Integrated Solution, Chris21, 2011, from <http://www.frontiersoftware.com/products/chris21-integrated-human-resource-management-payroll>

A Comparative study of Arabic handwritten characters invariant feature

Hamdi Hassen
Mir@cl Lab, FSEGS
University of Sfax
BP 1088, 3018Sfax, Tunisia
(216) 74 278 777
hassen2006@yahoo.fr

Maher Khemakhem
Mir@cl Lab, FSEGS
University of Sfax
BP 1088, 3018 Sfax, Tunisia
(216) 74 278 777
maher.khemakhem@fsegs.rnu.tn

Abstract—this paper is practically interested in the unchangeable feature of Arabic handwritten character. It presents results of comparative study achieved on certain features extraction techniques of handwritten character, based on Hough transform, Fourier transform, Wavelet transform and Gabor Filter. Obtained results show that Hough Transform and Gabor filter are insensitive to the rotation and translation, Fourier Transform is sensitive to the rotation but insensitive to the translation, in contrast to Hough Transform and Gabor filter, Wavelets Transform is sensitive to the rotation as well as to the translation.

Keywords- component ; Arabic handwritten character; invariant feature; Hough transform; Fourier transform; Wavelet transform; Gabor Filter.

I. INTRODUCTION

In order to improve the recognition rate of Arabic handwritten character, several techniques based on geometric correction or mathematic transform have been tested and applied on the input image of such writing. The purpose of all these techniques is to attempt to find the invariant features of Arabic handwritten characters. Unfortunately, all these attempts failed because, researchers didn't try to take advantages of each one of these techniques and then try to build an approach based on the cooperation of some (or all) of them.

Consequently, we show in this paper the behavior of some well-known mathematic transforms to decide in a next step on which of them can be used especially in a future Arabic OCR system based on the cooperation of several approaches and techniques.

The remainder of this paper is organized as follows: the first section presents a detailed description of the bench of test used (retained system) and the corresponding features. The second section, gives an assessment of this system, whereas the last section compares and analyses the obtained results.

II. THE RETAINED SYSTEM

Due to the nature of handwriting with its high degree of variability and imprecision obtaining these features, is a difficult task [1]. Feature extraction methods are based on 3

types of features: statistical, structural and global transformations and moments.

Several mathematic transform have been adapted to the Arabic character as well as for the modeling of the different morphological variations of the characters and for the resolution of the difficult problems of the segmentation.

A. Hough Transform

The Hough transform is a very general technique for feature detection. In the present context, we will use it for the detection of straight lines as contour descriptors in edge point arrays [2].

The Hough transformation (HT) is proposed since 1962 in US by P. V. C. Hough as the patent, because it may carry on the effective recognition to the shape, and parallel realizes, moreover is insensitive to the noise, thus obtained the enormous attention. In the practical application, mainly uses in recognizing curve[3].

Hough transform is a methodology which consists of the following three steps; the first one includes preprocessing for image enhancement, connected component extraction and average character height estimation. In the second step, a block-based Hough transform is used for the detection of potential text lines while a third step is used to correct possible false alarms [4].

1) Preprocessing

First, an adaptive binarization and image enhancement technique is applied [5]. Then, the connected components of the binary image are extracted [6], and for every connected component, the bounding box coordinates and the corresponding area are calculated [7]. Finally, the average character height AH for the document image is calculated [6]. We assume that the average character height equals to the average character width AW .

2) Hough Transform Mapping

In this stage, the Hough transform takes into consideration a subset (denoted as "subset 1" in Figure 1) of the connected components of the image. This subset is chosen for the following reasons: (i) it is required to ensure that components

which appear in more than one line will not vote in the Hough domain; (ii) components, such as vowel accents, which have a small size must be rejected from this stage because they can cause a false text line detection by connecting all the vowel above to the core text line.

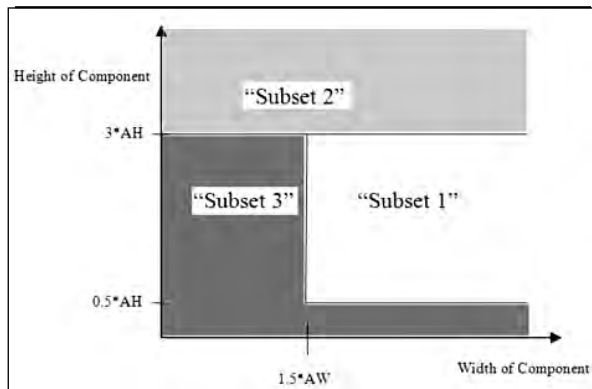


Figure 1. The connected component space divided into 3 subsets denoted as “Subset 1”, “Subset 2” and “Subset 3”.

The spatial domain for “subset 1” includes all components with a size identified by the following constraints [6]:

$$0.5*AH < H < 3* AH$$

$$1.5*AW < W < \Omega$$

Where H , W denote the component’s height and width, respectively, and AH , AW denote the average character height and the average character width, respectively.

After this partitioning stage, we calculate the gravity center of the connected component contained in each block which is then used in the voting procedure of the Hough transform.

The Hough transform is a line to point transformation from the Cartesian space to the polar coordinate space. Since a line in the Cartesian coordinate space is described by the equation [8]:

$$P = X \cos(\theta) + Y \sin(\theta) \quad [1]$$

To construct the Hough domain the resolution along θ direction was set to 1 degree letting θ take values in the range 85 to 95 degrees and the resolution along p direction was set to $0.2*AH$ [9]. After the computation of the accumulator array we proceed to the following procedure: we detect the cell (p_i, θ_i) having the maximum contribution and we assign to the text line (p_i, θ_i) . all points that vote in the area $(p_i - 5, \theta_i) .. (p_i + 5, \theta_i)$

3) 3) Post processing

The previous stage may result in more than one line assigned in the Hough domain that corresponds to a single text line (see Figure 1).

This correspondence is determined by calculating the distance between the corresponding crossing points of the lines with the document middle vertical line [10].

If the distance is less than the average distance of adjacent lines then all connected components which correspond to these lines are assigned to the same text line label. In this step we determinate the subset 2 and 3 (figure 1) “Subset 2” includes the components whose heights exceed 3 times the average height [6] (see Figure 1). These ‘large’ components may belong to more than one text line. This situation may appear when an ascender of one line meets a descender of an adjacent line. “Subset 3” includes all the components that do not fall into the previous two categories. Components of “Subset 3” are usually punctuation marks or elongations.

B. Wavelet Transform

Wavelet Transform is the method of decomposition that has gained a great deal of popularity in recent years. It seeks to represent a signal with good resolution in both time and frequency, by using basis functions called wavelets.

This algorithm consists in the following: for every level of decomposition, we make the extraction of the average of vertical, horizontal, right and left, directional and global features. Then, we decompose the picture of size $512 * 512$ in blocks of size $4*4$ stamps to have in total $16 * 5 = 80$ features then one calculates the black pixel density for every block [11]. The related components that have an elevated frequency density are filtered by 2D wavelets.

Mathematically, the process of Wavelets Transform is represented by the following equation:

$$F(w) = \int_{-\infty}^{+\infty} f(t) e^{-j\omega t} dt \quad [3]$$

With

$F(w)$: Wavelet Transform result,

$f(t)$: Input picture,

W : Sinusoide of frequency.

Wavelet transform are of different types: Haar Wavelet Transform, Symlet Wavelet Transform, Daubechies (db4) Wavelet transform, ..[12]

For every type, we can have two types of wavelet transform: the Continuous Wavelet Transform (CWT) and the Discrete Wavelet Transform (DWT) [13]. Moreover the wavelet is classified in several families Haar, Daubechies, Symlets, Coiflets...

And For every category, we can have several levels of coefficients [14] Daubechies (db3), Symlets (sym5), Coiflet2,... of which every category uses a coefficient that gives an invariant average in several representation to different resolution.

C. Gabor Filter

Gabor Filter can capture salient visual properties such as spatial localization, selectivity orientation, and spatial frequency characteristics [15][16]. We have chosen Gabor features to represent the face image considering these excellent capacities [17]. Gabor filters are defined as follows [18]:

$$G(x, y, \theta, f) = \exp \left[\frac{-1}{2} \left(\frac{x'}{sx'} \right)^2 + \left(\frac{y'}{sy'} \right)^2 \right] * \cos(2\pi * f * x') \quad [2]$$

With

$$x' = x * \cos(\theta) + y * \sin(\theta)$$

$$y' = y * \cos(\theta) - x * \sin(\theta);$$

I: Input picture

Sx & Sy: The Variances of the x and y axes respectively

f: the frequency of the sinisouide function

θ : The orientation of Gabor filter

G: the output of Gabor filter.

The feature vector used by the OCR system consists in the combination of several parameters such as θ , f and the other parameters [18].

The rotation of the characters in the X and Y axes by angle θ , results in a vector of feature with an orientation θ .

The value $\theta = \frac{2\pi}{m} * (k - 1)$ with $k = 1 \dots m$ and m is the number of orientations

The position of a character in a N*N picture is taken into consideration in order to get a result of invariant feature vector to the rotation and translation [19].

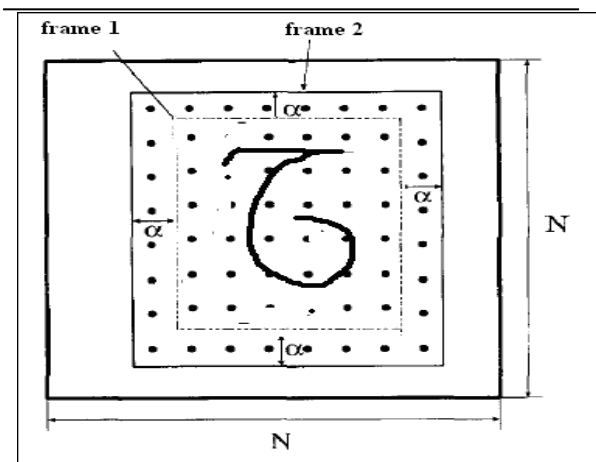


Figure2. Position of the character «Haa» in a picture of [N*N] size

D. Fourier Transform

The Fourier Transform (FT) of the contour of the image is calculated. Since the first n coefficients of the FT can be used in order to reconstruct the contour, then these n coefficients are considered to be a n-dimesional feature vector that represents the character

The recognition of the writing by Fourier Transform is operated previously on the contour of the character [20]. At the first stage, we start with the detection of the contour. In the second stage, the code of Freeman of the contour is generated on which one operates the calculation of Fourier Transformed.

The figure below shows the stages of character recognition by Fourier Transform.

The method Fourier Transform requires the modelling of the function of the contour. The contour of a character is described by a sequence of the codes of Freeman characterized by [20]:

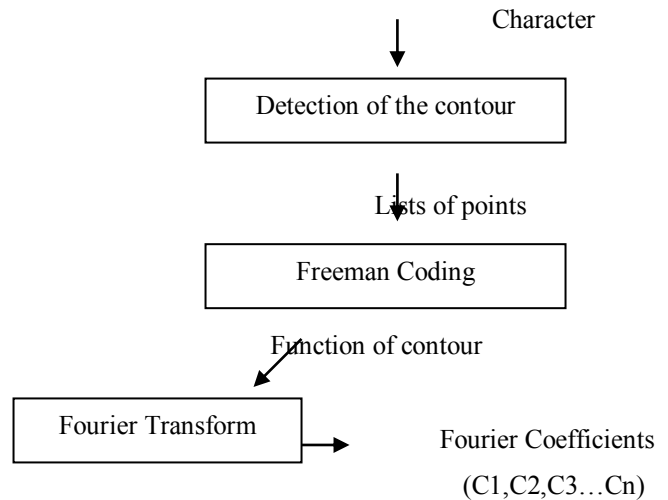


Figure 3. The stages of Fourier Transform

The values: between 0 and 7,

The direction: of 0 to 2π per step of $\pi/4$ while following the opposite direction of the needle of a watch,

The length: 1 or $\sqrt{2}$ according to the parity of the codes.

From this description, we can generate a periodic temporal function on which Fourier Transform can be calculated.

Mathematically, the process of Fourier Transform is represented by the following equation:

$$X_N(K) = A_0 + \sum_{n=1}^N a_n \cos \frac{2\pi nK}{N} + b_n \sin \frac{2\pi nK}{N} \quad [4]$$

$$Y_N(K) = C_0 + \sum_{n=1}^N c_n \cos \frac{2\pi nK}{N} + d_n \sin \frac{2\pi nK}{N} \quad [5]$$

Where

K: the kème points of the contour,

N: is the necessary number in the approximation of the contour by the coefficients of Fourier

a_n, b_n, c_n et d_n : The coefficients of Fourier corresponding to the harmonic n.

a_0 et c_0 : The continuous components that correspond to the initial points where the frequency is equal to 0.

III. EVALUATION OF THE RETAINED SYSTEM.

A. Implementation

For some experimental convenient, the MATLAB Version 7.4.0 (R2007a) (MATrix LABoratory) has been used.

B. The IFN/ENIT Handwritten Character Database

We have used the well known IFN/ENIT corpus data base formed of handwritten Tunisian town's names.

The figure 4 presents a sample basis of the IFN/ENIT.

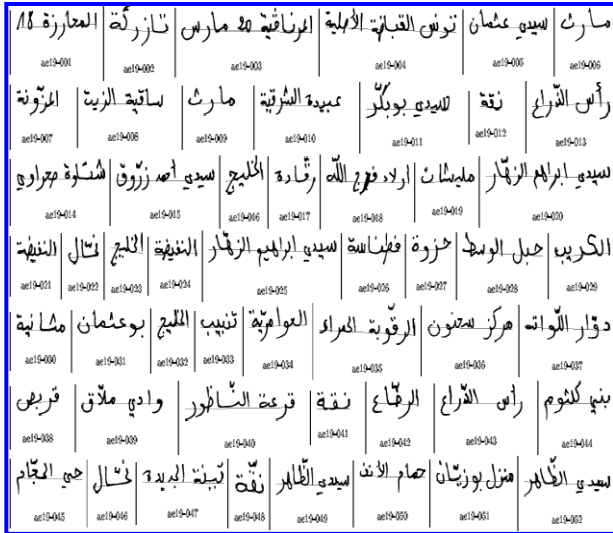


Figure 4. Some names of Arabic Tunisian towns of the IFN/ENIT

These names have undergone of the segmentations, the final version of the retained basis includes the different handwritten Arabic characters in the different positions (initial, median, isolated).

We note that the set of characters, the signs of punctuation as well as the numbers are written by different writers. (The basis of training includes 692 pictures of letters whereas the basis of test includes 1674 which 692 those of the training and 982 new pictures of characters to be recognized :)

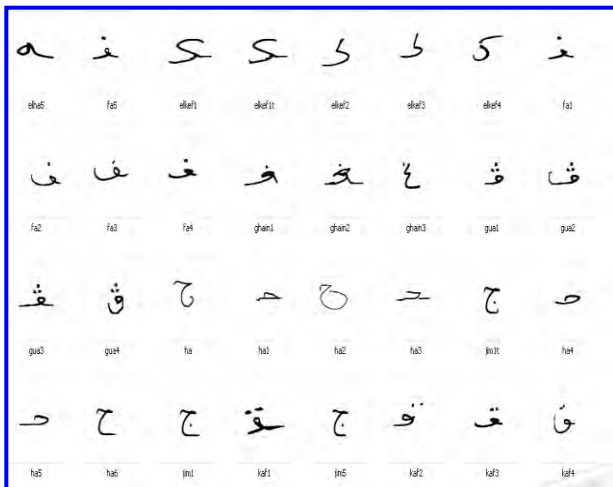


Figure 5. Some handwritten Arabic characters pictures

C. The classification by Euclidean Distance

There are several methods of classification such as k-Nearest Neighbour (k-NN) , Bayes Classifier, Neural Networks (NN), Hidden Markov Models (HMM), Support Vector Machines (SVM), etc [21].

There is no such thing as the “best classifier”. The use of classifier depends on many factors, such as available training set, number of free parameters etc.

In our case, we used Euclidean Minimum Distance Classifier (EMDC).

The consistent rule is relative to the decision of the nearest neighbor's; the idea is extremely simple the character feature vector x_i is compared with the vectors y_i describing the character in the Base Set to search for the nearest neighbour of the tested character.

The general equation of the Euclidean distance is the following:

$$d(x, y) = \sqrt{\sum_{i=1}^N (x_i - y_i)^2} \quad [6]$$

D. Test and evaluation

The implementation of our system requires the definition of a set of parameters:

Wavelets Daubechies (db) with a level of decomposition equal to 3 because they are used intensively and they present interesting properties.

FFT2: Fast Fourier Transform level 2 is the algorithm of Fourier used in our implementation because the calculation of the coefficients of discrete Fourier is less expensive.

The implementation of Hough Transform doesn't require a particular parameter; we simply program the different stages of this method.

The implementation of the Gabor Filters method requires the research of the most combinations of a set of parameters ($\lambda, \sigma_x, \sigma_y, m$) that maximizes the recognition rate.
 $m = 4 ; \lambda = 1; \sigma_x = 2; \sigma_y = 1;$

The results of our study are reported on the following figures.

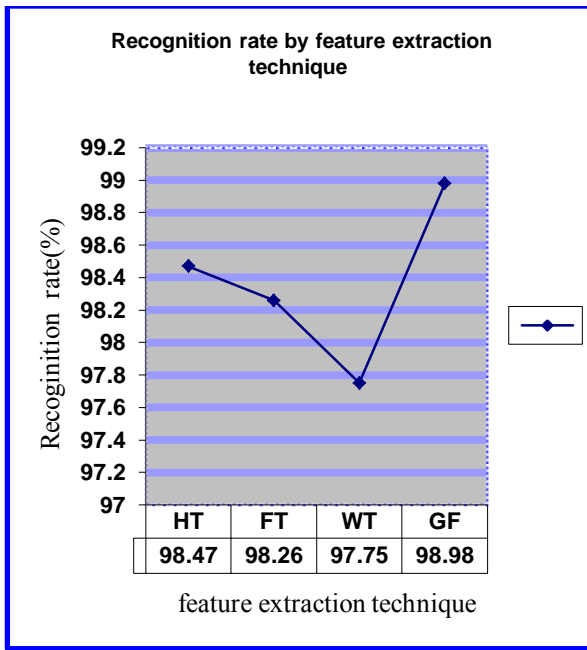


Figure 6. Recognition Rate by feature extraction technique

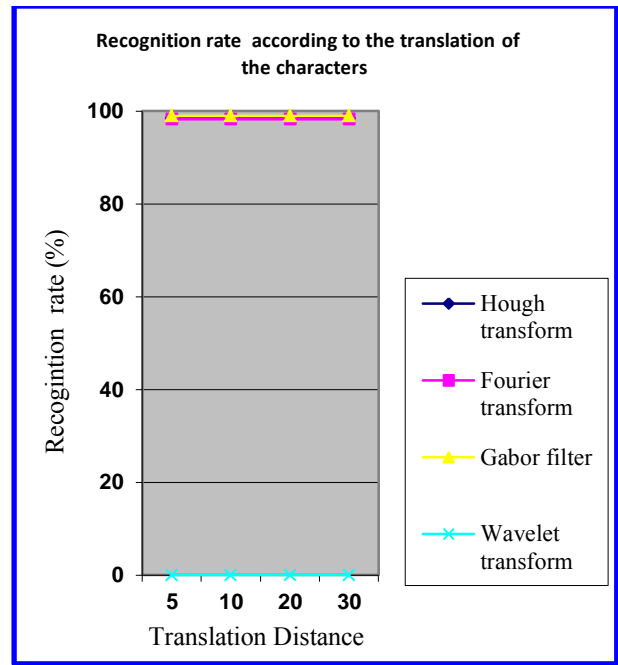


Figure 8. Recognition rate of according to the translation of the characters

After extensive experiments, we tried to calculate the response time of our recognition system.

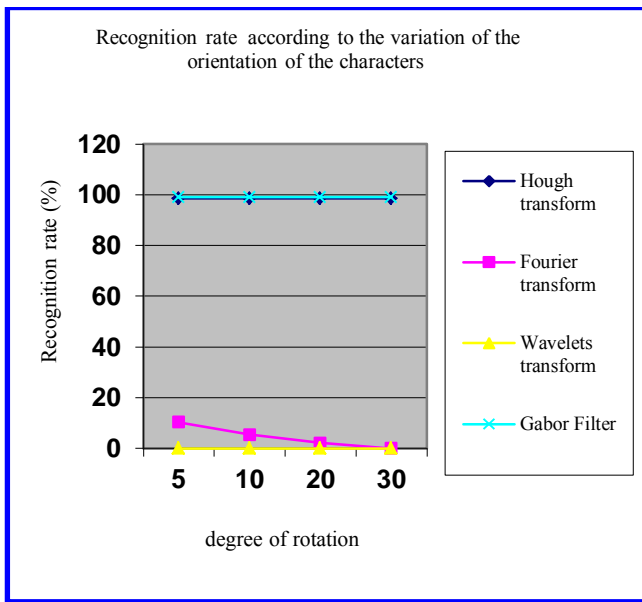


Figure 7. Recognition rate according to the variation of the orientation of the characters

Similarly, the transfer of the characters according to the x axis or y coordinates gives interesting results with Gabor, Hough and Fourier transforms.

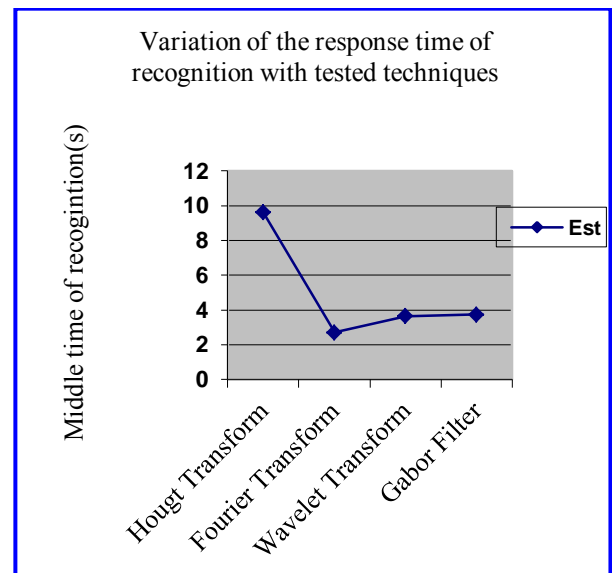


Figure 9. Variation of the response time of recognition with tested techniques

We have evaluated also the size characters (written) variation with the four extraction techniques of our system. Obtained results are reported in the figure below .

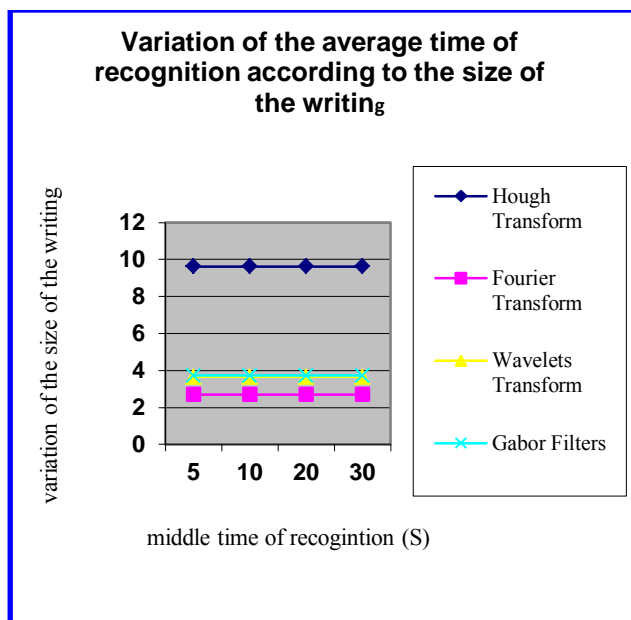


Figure 10. Variation of the average time of recognition according to the size of the writing

E. Interpretation:

The previous results show that:

The four feature extraction techniques can give interesting recognition rates depending to the initial conditions of the Arabic writing, Hough Transform and Gabor filter are insensible to the rotation and translation,

Hough Transform and Gabor filter are indeed efficient and strong tools for the detection of remarkable shapes in a picture,

Hough Transform, was very slow because of the corresponding complexity,



Fourier Transform is sensible to the rotation but insensible to the translation,



In contrast to Hough Transform and Gabor filter, Wavelets Transform is sensitive to the rotation as well as to the translation,



The writing size variation doesn't have an influence on the recognition average time for the four studied extraction techniques (average time is invariant)

IV. SURVEY OF SOME DETECTED ERRORS.



The most common detected errors are the following



Confusion between the characters «gua »  and « tha »  (Fourier transform),



Confusion between the characters « ya »  and « hamza »  (Gabor Filter),

Confusion between the characters « kaf »  and « fa »  : the two diacritical points of the character

" kaff " are confounded in one only. (Hough transform),

Confusion between the characters « elha »  and « mim »  (Wavelets Transform).

The two characters (mim)  and (zed)  have the same Euclidian distance (Wavelet transform),

The two characters (lam)  and (del)  have the same Euclidian distance (Hough transform)

V. CONCLUSION AND FUTURE WORK

We present in this paper a comparative study of different features extraction techniques based on Hough Transform, Wavelet Transform, Fourier Transform and Gabor filters.

The obtained results provide interesting idea about the studied techniques.

These results show that Hough Transform and Gabor filter are insensible to the rotation and translation, Fourier Transform is sensible to the rotation but insensible to the translation, in contrast to Hough Transform and Gabor filter, Wavelets Transform is sensitive to the rotation as well as to the translation

The considered perspectives to make evolve the present work are numerous: we can note the evolution of other classification is one of them, exploiting new features to improve the current performance. The integration and cooperation of some complementary approaches that can lead to an effective Arabic OCR is also another way of investigation.

The digitization of the national cultural heritage is another future project or work

The project is expected to connect with vanguard digital libraries such as Google, and to digitize many books, periodicals and manuscripts.

Our project would make access to data much easier for researchers and students in all regions of the country.

We need a technology that offers a number of benefits, such as the ability to store and retrieve large amounts of data in any location at any time (Mobility of the users: dynamic environment).

ACKNOWLEDGMENT

The authors wish to thank the reviewers for their fruitful Comments. The authors also wish to acknowledge the members of the miracl laboratory, sfax, tunisia

REFERENCES

- [1] Margner V., Abed H. E. Arabic Handwriting Recognition Competition, ICDAR, 2007, pp. 1274-1278.
- [2] Charles V.StewarsComputer Vision, The Hough Transform, February 24, 2011
- [3] Jinxue Sui, Li Yang, Zhen Hua International Journal of Digital Content Technology and its Applications. Volume 5, Number 3, March 2011
- [4] Khalid Saeed , Majida Albakoor "Region growing based segmentation algorithm for typewritten and handwritten text recognition"2009.
- [5] Fu Chang, Chun-jen Chen, Chi-Jen Lu, « A linear-time component- Labelling Algorithm using contour tracing technique », computer vision and image understanding, vol. 93. No.2, fabruary 2004, pp.206-220

- [6] B. Gatos, I. Pratikakis and S. J. Perantonis, "Adaptive Degraded Document Image Binarization", Pattern Recognition, Vol. 39, pp. 317-327, 2006.
- [7] B.Gatos,T. Konidaris, K.Ntzios, I. Pratikakis and S. J. Perantonis,"A Segmentation-free Approach for Keyword Search in Historical Typewritten Document" 8th International Conference on Document Analysis and Recognition (ICDAR'05), Seoul, Korea, August 2005.
- [8] P.Etyngier, N.Paragios, J.-Y. Audibert., and R. Keriven, Radon/hough space for pose estimation, Rapport de recherche 06-22, Centre d'Enseignement et de Recherche en Technologies de l'Information et Systèmes, Ecole Nationale des Ponts et Chaussées, Janvier 2006.
- [9] Z. Shi, S. Setlur, and V. Govindaraju, "Text Extraction from Gray Scale Historical Document Images Using Adaptive Local Connectivity Map", Eighth International Conference on Document Analysis and Recognition, Seoul, Korea, 2005, pp. 794-798.
- [10] S. Nicolas, T. Paquet, L. Heutte, "Text Line Segmentation in Handwritten Document Using a Production System", Proceedings of the 9th IWFHR, Tokyo, Japan, 2004, pp. 245-250.
- [11] Zheng, L., Hassin, A., Tang, X., "A New Algorithm for Machine Printed Arabic Character Segmentation." Pattern Recognition Letters 25, 2004.
- [12] Rohit Arora et al. / International Journal of Engineering Science and Technology (IJEST) Vol. 3 No. 4 Apr 2011
- [13] Ze-Nian & Drew "Fundamentals of Multimedia, Pearson Prentice Hall", 2004.
- [14] Saeed M., Karim F, & Hamidreza R. (2004a) Feature Comparison between Fractal Codes and Wavelet Transform in Handwritten Alphanumeric Recognition Using SVM Classifier. Proc. IEEE 7th International Conference on Pattern Recognition (ICPR'04).
- [15] P.N.Belhumeur, J.P.Hespanha etc. Eigenfaces, Eigenfaces vs Fisherfaces: «recognition using class specific linear projection. » IEEE Trans. Pattern Analysis and Machine Intelligence vol.20, no.7, pp.711-720, 1997
- [16] S.Shan, « Study on some key issues in face recognition», Ph.D thesis, CAS, 2004.
- [17] A. AL-DMOUR, R. ABU ZITAR, "Arabic writer identification based on hybrid spectral-statistical measures", Journal of Experimental & Theoretical Artificial Intelligence, Volume 19, 2007, pp. 307-332.
- [18] Kamarainen, J., «Feature extraction using Gabor filters», PhD thesis, Lappeenranta University of Technology, 2003.
- [19] Taavi Aalto Atte Kilpelä Jukka Lankinen E. Antero Tammi «Gabor filtering Case: Feature extraction» 22.9.2006.
- [20] M.Szmulo, «Boundary normalization for recognition of non touching non-degraded characters », ICDAR, IEEE, 1997, pp 463-466.
- [21] C. Huang, S. Srihari, Word segmentation of off-line handwritten documents, in: Proceedings of the Document Recognition and Retrieval (DRR) XV, IST/SPIE Annual Symposium, San Jose, CA, USA, January 2008.
- [22] A project report: RECOGNITION OF HANWRITING ON POSTAL LETTERS AND PARCELS BY DANIEL CHUA.
- [23] A Block-Based Hough Transform Mapping for Text Line Detection in Handwritten Documents by G. Louloudis, B. Gatos, I. Pratikakis, K. Halatsis.

AUTHORS PROFILE



Hassen Hamdi received in 2008 his Master's Degree in Computer Science from the University of Sfax, Tunisia. He is currently a Ph.D student at the University of Sfax. His research interests include distributed systems, performance analysis, Networks security and pattern recognition.



Maher Khemakhem received his Master of Science, his Ph.D. and Habilitation degrees from the University of Paris 11 (Orsay), France respectively in 1984, 1987 and the University of Sfax, Tunisia in 2008. He is currently Associate Professor in Computer Science at the Higher Institute of Management at the University of Sousse, Tunisia. His research interests include distributed systems, performance analysis, Networks security and pattern recognition.

Pattern Discovery Using Association Rules

Ms Kiruthika M, Mr Rahul Jadhav
Associate Prof., Computer dept.
Fr CRIT, Vashi,
Navi Mumbai, India

Ms Rashmi J
Lecturer, IT dept.
FrCRIT, Vashi
Navi Mumbai, India

Ms Dipa Dixit
Assistant Prof., IT dept.
Fr CRIT, Vashi,
Navi Mumbai, India

Ms Anjali Nehete, Ms Trupti Khodkar
Fr CRIT, Vashi,
Navi Mumbai, India

Abstract— The explosive growth of Internet has given rise to many websites which maintain large amount of user information. To utilize this information, identifying usage pattern of users is very important. Web usage mining is one of the processes of finding out this usage pattern and has many practical applications. Our paper discusses how association rules can be used to discover patterns in web usage mining. Our discussion starts with preprocessing of the given weblog, followed by clustering them and finding association rules. These rules provide knowledge that helps to improve website design, in advertising, web personalization etc.

Keywords- Weblogs; Pattern discovery; Association rules.

I. INTRODUCTION

Association rule is one of the data mining tasks which can be used to uncover relationship among data. Association rule identifies specific association among data and its techniques are generally applied to a set of transactions in a database. Since, amount of data handled is extremely large, current association rule techniques are trying to prune the search space according to support count.

Rules discovery finds common rules in the format $A \rightarrow B$, meaning that, when page A is visited in a transaction, page B will also be visited in the same transaction. These rules may have different values of the confidence and support [1].

Confidence is the percentage between the number of transactions containing both items of the rule and the number of transactions containing just the antecedent. Support is the percentage of transactions in the rule is true.

In the context of Web Usage Mining, association rules refers to set of pages which are accessed together with a minimum support value which can help in organizing Web space efficiently.

For example: Consider if 70% of the users who accessed `get/programs/courses/x.asp` also accessed `get/programs/courses/y.asp`, but only 30% of those who accessed `get/programs/courses` accessed

`get/programs/courses/y.asp`, then it shows that some information in `x.asp` is making the clients access `y.asp`.

This inference helps the designers to decide on designing a link between the above two pages. The task of association rule mining has received a great deal of attention. Association rule mining is still one of the most popular pattern-discovery methods in KDD.

Hence, we would like to use association rules for pattern discovery analysis of Web Server Logs.

A. Web Server Log

Web Servers are used to record user interactions whenever any request for resources are received.

A server log is a log file automatically created and maintains a history of page requests. Information about the request, including client IP address, request date/time, page requested, HTTP code, bytes served, user agent, and referrer are typically added. These data can be combined into a single file, or separated into distinct logs, such as an access log, error log, or referrer log. However, server logs typically do not collect user-specific information [2].

But to understand the user behavior, analysis of these weblogs is a must. This analysis can help in understanding the user access patterns and can lead to grouping of resource providers, restructuring of websites, pinpointing effective advertising locations, targeting specific users for specific advertisements.

Unprocessed log are shown below:

```
#Fields: date time c-ip cs-username s-sitename s-  
computername s-ip s-port cs-method cs-uri-stem cs-uri-  
query sc-status time-taken cs-version cs-host cs(User-  
Agent) cs(Referer)  
2002-04-01 00:00:10 1cust62.tnt40.chi5.da.uu.net -  
w3svc3 bach bach.cs.depaul.edu 80 get  
/courses/syllabus.asp course=323-21-  
603&q=3&y=2002&id=671 200 156 http/1.1  
www.cs.depaul.edu
```

mozilla/4.0+(compatible;+msie+5.5;+windows+98;+win
+9x+4.90;+msn+6.1;+msnbnmsft;+msnmen-us;+msnc21)
http://www.cs.depaul.edu/courses/syllabilist.asp
depaul.edu/courses/syllabilist.asp

2002-04-01 00:00:26 ac9781e5.ipt.aol.com - w3svc3
bach bach.cs.depaul.edu 80 get /advising/default.asp -
200 16 http/1.1 www.cs.depaul.edu
mozilla/4.0+(compatible;+msie+5.0;+msnia;+windows+
98;+digext)
http://www.cs.depaul.edu/news/news.asp?theid=573

2002-04-01 00:00:29 alpha1.csd.uwm.edu - w3svc3
bach bach.cs.depaul.edu 80 get /default.asp - 302 0
http/1.1www.cs.depaul.edu
mozilla/4.0+(compatible;+msie+6.0;+msn+2.5;+window
s+98;+luc+user) -

A sample log file converted into database is shown below
in Table I.

II. SCOPE AND APPLICATIONS

The user access log has very significant information about a
Web server. A Web server access log contains a complete
history of webpages accessed by clients. By analyzing these
logs, it is possible to discover various kinds of knowledge,
which can be applied to improve the performance of Web
services.

Web usage mining has several applications and is used in
the following areas:

- 1) It offers users the ability to analyze massive volume of
click stream or click flow data.
- 2) Personalization for user can be achieved by keeping track
of previously accessed pages which can be used to
identify the typical browsing behavior of a user and
subsequently to predict desired pages.

- 3) By determining access behavior of users, needed links can
be identified to improve the overall performance of future
accesses.

Web usage patterns are used to gather business intelligence
to improve customer attraction, customer retention, sales,
marketing, and advertisements cross sales. Web usage mining
is used in e-Learning, e-Business, e-Commerce, e-Newspapers,
e-Government and Digital Libraries.

III. PROPOSED SYSTEM

We would like to propose a system which would discover
interesting patterns in these weblogs. Weblogs has information
about accesses to various Web pages within the Web space
associated with a particular server.

In case of Web transactions, association rules capture
relationships among pageviews based on navigation patterns of
users.

A. Steps involved in the proposed system

Our proposed system would involve the following steps:

- 1) The input is a set of Weblogs for which we have to find
association rules. We have chosen University Web server
logs from www.cs.depaul.edu site
- 2) The server logs contain entries that are redundant or
irrelevant for data mining tasks.
- 3) The Data cleaning process will select a subset of fields
that are relevant for the task.
- 4) These selected attributes are then stored into a database.
- 5) Using a simple clustering approach these entries are
divided into clusters or segmented.
- 6) Now, association rule mining is applied on these clusters,
to obtain association rules having minimum support and
confidence.
- 7) As a result of association rule mining, interesting patterns
can be discovered and client's web usage can be
evaluated.

TABLE I: A SAMPLE LOG FILE IN TABLE FORMAT.

T no	Client IP	Date time	Method	Server IP	Port	URI Stem
0	202.185.122.151	11/23/200 3 4:00:01PM	GET	202.190.126.85	80	/index.asp
1	202.185.122.151	11/23/200 3 4:00:08 PM	GET	n202.190.126.85	80	/index.asp
2	210.186.180.199	11/23/200 3 4:00:10 PM	GET	202.190.126.85	80	/index.asp
3	210.186.180.199	11/23/200 3 4:00:13 PM	GET	202.190.126.85	80	/tutor/include/style03.css
4	210.186.180.199	11/23/200 3 4:00:13 PM	GET	202.190.126.85	80	/tutor/include/detectBrowser_ cookie.js

IV. DESIGN

A. Flow Diagram:

The flowchart for pattern discovery using association rules is given in fig 1.

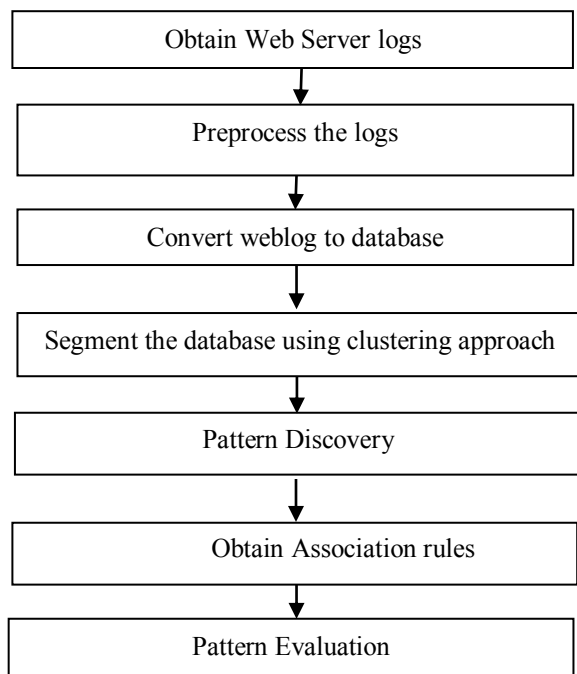


Figure 1 Flow diagram for pattern discovery of weblogs

Each of these blocks is explained in detail as follows:

1) Obtain Web Server logs:

Web server log is a file which is created and maintained by the webserver. We are analyzing the log file of the site: www.cs.depaul.edu. It is a text file. The file follows the extended log file format.

2) Preprocessing the logs:

The weblog created by the webserver contains details of all requests. It contains lot of irrelevant, incomplete data. Preprocessing involves removing such data.

3) Conversion of log file to database:

The weblog cannot be directly used for data mining. The dataset is converted to a database. This involves creating a database and then importing the log file to the MySQL database table.

4) Segmenting the database:

In this step, the database is segmented into clusters depending on the support count. After this a number of small clusters are obtained. Depending on the need, these clusters can be analyzed. Clustering web usage data allows the Web master to identify groups of users with similar behaviors for which personalized versions of the Web site may be created.

5) Pattern Discovery:

The next step is pattern discovery. Once the clusters are formed they are studied to recognize patterns within the entries of the clusters.

6) Association rules:

Association rules show relationship among different items. In case of Web mining, an example of an association rule is the correlation among accesses to various web pages on a server by a given client. Such association rules are obtained in this step

7) Pattern Evaluation:

The association rules obtained in the earlier step help in establishing relationships among data items. These association rules are evaluated to understand the information they provide. The interpretations of the rules provide useful knowledge.

B. Implementation

The following diagrams illustrate the steps of implementation.



Figure 2 Welcome screen of weblog analyzer

Step 1: Weblog of University website hosted on a web server were obtained from www.cs.depaul.edu. There are 5061 records. The following figure3 shows the unprocessed weblog file.

```
#Software: Microsoft Internet Information Services 5.0
#Version: 1.0
#Date: 2002-04-01 00:00:00
#Fields: date time c-ip cs-username s-sitename s-computername s-ip s-port cs-method cs-uri-stem cs-uri-query sc-status t
2002-04-01 00:00:10 10.0.0.10 10.0.0.10 cs-get /courses/syllabus.asp course=32
2002-04-01 00:00:26 a:9781e5.1pt.aol.com - w3svc3 bach bach.cs.depaul.edu 80 get /advising/default.asp - 200 16 http://1.1
2002-04-01 00:00:29 a1ghal.csd.uwm.edu - w3svc3 bach bach.cs.depaul.edu 80 get /default.asp - 302 0 http://1.1 www.cs.depa
2002-04-01 00:00:30 w010.2064221069.chi-ft1.ds1.cnc.net - w3svc3 bach bach.cs.depaul.edu 80 get /default.asp - 200 94
2002-04-01 00:00:30 w010.2064221069.chi-ft1.ds1.cnc.net - w3svc3 bach bach.cs.depaul.edu 80 get /default.asp - 302 0 http
2002-04-01 00:00:30 a1ghal.csd.uwm.edu - w3svc3 bach bach.cs.depaul.edu 80 get /news/default.asp - 200 62 http://1.1 www.c
2002-04-01 00:00:30 w010.2064221069.chi-ft1.ds1.cnc.net - w3svc3 bach bach.cs.depaul.edu 80 get /news/default.asp - 200 6
2002-04-01 00:00:32 a:9781e5.1pt.aol.com - w3svc3 bach bach.cs.depaul.edu 80 get /resources/uo_scholarships.asp section=
2002-04-01 00:00:34 chf-f111-202.rasserver.net - w3svc3 bach bach.cs.depaul.edu 80 get /courses/syllabus.asp course=468-
2002-04-01 00:00:35 12-250-96-248.c1ent.attol.com - w3svc3 bach bach.cs.depaul.edu 80 get /courses/syllabus.asp - 2
2002-04-01 00:00:36 w010.2064221069.chi-ft1.ds1.cnc.net - w3svc3 bach bach.cs.depaul.edu 80 get /programs/default.asp - 2
2002-04-01 00:00:40 a:9781e5.1pt.aol.com - w3svc3 bach bach.cs.depaul.edu 80 get /advising/nf_scholarships.asp - 200 62
2002-04-01 00:00:44 w010.2064221069.chi-ft1.ds1.cnc.net - w3svc3 bach bach.cs.depaul.edu 80 get /courses/2002/grades2002
2002-04-01 00:01:00 66-79-37-44.coastalnow.net - w3svc3 bach bach.cs.depaul.edu 80 get /resources/gae_guide.asp |-|0404
2002-04-01 00:01:00 66-79-37-44.coastalnow.net - w3svc3 bach bach.cs.depaul.edu 80 get /shared/404.asp 404:http://www.cs
2002-04-01 00:01:07 w010.2064221069.chi-ft1.ds1.cnc.net - w3svc3 bach bach.cs.depaul.edu 80 get /programs/courses.asp dep
2002-04-01 00:01:09 chf-f111-202.rasserver.net - w3svc3 bach bach.cs.depaul.edu 80 get /courses/syllabus.asp - 200 71
2002-04-01 00:01:14 w010.2064221069.chi-ft1.ds1.cnc.net - w3svc3 bach bach.cs.depaul.edu 80 get /people/facultyinfo.asp i
2002-04-01 00:01:15 ac90e8a.1pt.aol.com - w3svc3 bach bach.cs.depaul.edu 443 post /ctf/advising/display.asp - 200 1637
2002-04-01 00:01:20 chf-f111-202.rasserver.net - w3svc3 bach bach.cs.depaul.edu 80 get /courses/syllabus.asp course=468-
2002-04-01 00:01:31 10.0.0.10 10.0.0.10 cs-get /courses/syllabus.asp course=31
2002-04-01 00:01:36 chf-f111-202.rasserver.net - w3svc3 bach bach.cs.depaul.edu 80 get /courses/syllabus.asp - 200 71
2002-04-01 00:01:43 ac90e8a.1pt.aol.com - w3svc3 bach bach.cs.depaul.edu 443 post /ctf/advising/display.asp - 302 2875
2002-04-01 00:01:43 ac90e8a.1pt.aol.com - w3svc3 bach bach.cs.depaul.edu 443 get /ctf/advising/display.asp - 200 47 htt
2002-04-01 00:01:47 12-250-96-248.c1ent.attol.com - w3svc3 bach bach.cs.depaul.edu 80 post /courses/syllabus.asp - 2
2002-04-01 00:01:53 12-249-142-45.c1ent.attol.com - w3svc3 bach bach.cs.depaul.edu 443 get /courses/syllabussearch.asp -
2002-04-01 00:01:58 chf-f111-202.rasserver.net - w3svc3 bach bach.cs.depaul.edu 80 get /courses/syllabus.asp course=354-
2002-04-01 00:02:01 12-250-96-248.c1ent.attol.com - w3svc3 bach bach.cs.depaul.edu 80 get /courses/syllabus.asp course=
2002-04-01 00:02:08 ac90e8a.1pt.aol.com - w3svc3 bach bach.cs.depaul.edu 443 get /ctf/advising/Includes/Faculty/account
7/07-04-01 00:02:01 rache-ent-ant11.annov.aol.com - w3svc3 bach bach.cs.depaul.edu 80 get /advising/default.asp - 200 47 ht
```

Figure 3 Unprocessed log file of www.cs.depaul.edu

Step 2: The next step is to convert the log file to database.

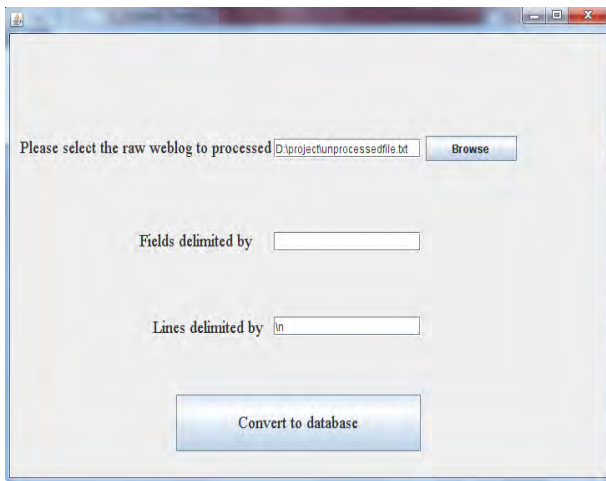


Figure 4 Selection of log file, specifying line and fields delimiters for the selected log file

The steps involved in the conversion of dataset to database are as follows:

- Log on to MySQL command line client.
- Create a table with all required attributes.
- Import the log files into database.

The MySQL commands are shown in figure below:

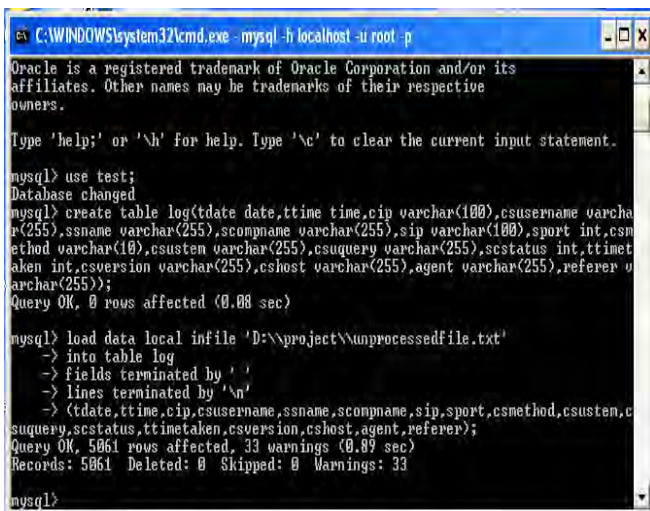


Figure 5 MySQL command prompt showing the commands used to create and load log database.

The database containing the entries of weblog is shown below in Fig 6

Step 3: The database has 5061 records. The count of entries for different IP addresses is obtained. There are entries having very low support count. Such entries need not be considered. The database is segmented into clusters having support count more than 20.

Step 4: The entries for IP addresses having support count greater than or equal to 30 are used for further analysis. There are 8 unique IP addresses having support count greater than or equal to 30.

These IP addresses are shown in fig 10.



Figure 6 The database of weblog entries

The following fig 7 shows the entries of IP addresses having support count greater than or equal to 20.

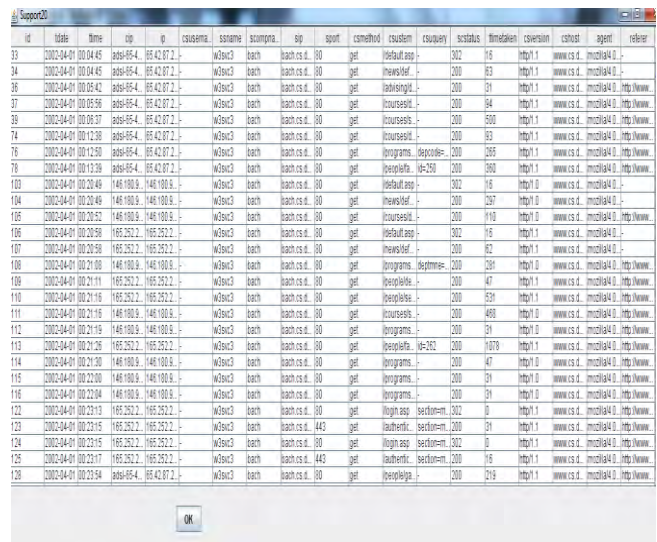


Figure 7 Weblog entries with IP addresses having support count greater than or equal to 20

The following fig 8 shows the entries of IP addresses having support count greater than or equal to 25.

id	date	time	ip	csystem	username	session	sp	sport	csmethod	csystem	csystem	session	csystem	agent	referer
93	2002-04-04	00:04:45	146.180.9.145	w3svc3	bach	bach.cs.	80	get	Default.asp	302	116	http://	www.cs.mccall.ca	-	
94	2002-04-04	00:04:45	146.180.9.145	w3svc3	bach	bach.cs.	80	get	Default.asp	302	116	http://	www.cs.mccall.ca	-	
95	2002-04-04	00:04:45	146.180.9.145	w3svc3	bach	bach.cs.	80	get	Default.asp	302	116	http://	www.cs.mccall.ca	-	
96	2002-04-04	00:05:56	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
97	2002-04-04	00:05:56	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
98	2002-04-04	00:06:37	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
99	2002-04-04	00:06:37	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
100	2002-04-04	00:07:28	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
101	2002-04-04	00:07:28	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
102	2002-04-04	00:07:28	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
103	2002-04-04	00:07:28	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
104	2002-04-04	00:07:28	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
105	2002-04-04	00:07:28	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
106	2002-04-04	00:07:28	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
107	2002-04-04	00:07:28	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
108	2002-04-04	00:07:28	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
109	2002-04-04	00:07:28	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
110	2002-04-04	00:07:28	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
111	2002-04-04	00:07:28	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
112	2002-04-04	00:07:28	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
113	2002-04-04	00:07:28	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
114	2002-04-04	00:07:28	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
115	2002-04-04	00:07:28	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
116	2002-04-04	00:07:28	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
117	2002-04-04	00:07:28	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
118	2002-04-04	00:07:28	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
119	2002-04-04	00:07:28	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
120	2002-04-04	00:07:28	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
121	2002-04-04	00:07:28	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
122	2002-04-04	00:07:28	146.180.9.145	w3svc3	bach	bach.cs.	80	get	courses	302	39	http://	www.cs.mccall.ca	http://www...	
123	2002-04-04	00:07:28	146.180.9.145	w3svc3	bach	bach.cs.	443	get	Default.asp	302	116	http://	www.cs.mccall.ca	-	
124	2002-04-04	00:07:28	146.180.9.145	w3svc3	bach	bach.cs.	443	get	Default.asp	302	116	http://	www.cs.mccall.ca	-	

Figure 8 Weblog entries with IP addresses having support count greater than or equal to 25

The following fig 9 shows the entries of IP addresses having support count greater than to equal to 30

id	date	time	ip	csystem	username	session	sp	sport	csmethod	csystem	csystem	session	csystem	agent	referer
106	2002-04-04	00:20:50	165.252.27.130	w3svc3	bach	bach.cs.	80	get	Default.asp	302	116	http://	www.cs.mccall.ca	-	
107	2002-04-04	00:20:56	165.252.27.130	w3svc3	bach	bach.cs.	80	get	Default.asp	302	116	http://	www.cs.mccall.ca	-	
108	2002-04-04	00:21:08	146.180.9.145	w3svc3	bach	bach.cs.	80	get	Default.asp	302	116	http://	www.cs.mccall.ca	-	
109	2002-04-04	00:21:11	165.252.27.130	w3svc3	bach	bach.cs.	80	get	Default.asp	302	116	http://	www.cs.mccall.ca	-	
110	2002-04-04	00:21:16	165.252.27.130	w3svc3	bach	bach.cs.	80	get	Default.asp	302	116	http://	www.cs.mccall.ca	-	
111	2002-04-04	00:21:16	146.180.9.145	w3svc3	bach	bach.cs.	80	get	Default.asp	302	116	http://	www.cs.mccall.ca	-	
112	2002-04-04	00:21:19	146.180.9.145	w3svc3	bach	bach.cs.	80	get	Default.asp	302	116	http://	www.cs.mccall.ca	-	
113	2002-04-04	00:21:26	165.252.27.130	w3svc3	bach	bach.cs.	80	get	Default.asp	302	116	http://	www.cs.mccall.ca	-	
114	2002-04-04	00:21:30	146.180.9.145	w3svc3	bach	bach.cs.	80	get	Default.asp	302	116	http://	www.cs.mccall.ca	-	
115	2002-04-04	00:22:00	146.180.9.145	w3svc3	bach	bach.cs.	80	get	Default.asp	302	116	http://	www.cs.mccall.ca	-	
116	2002-04-04	00:22:04	146.180.9.145	w3svc3	bach	bach.cs.	80	get	Default.asp	302	116	http://	www.cs.mccall.ca	-	
122	2002-04-04	00:23:13	165.252.27.130	w3svc3	bach	bach.cs.	80	get	Default.asp	302	116	http://	www.cs.mccall.ca	-	
123	2002-04-04	00:23:15	165.252.27.130	w3svc3	bach	bach.cs.	443	get	Default.asp	302	116	http://	www.cs.mccall.ca	-	
124	2002-04-04	00:23:15	165.252.27.130	w3svc3	bach	bach.cs.	443	get	Default.asp	302	116	http://	www.cs.mccall.ca	-	

Figure 9 Weblog entries with IP addresses having support count greater than or equal to 30

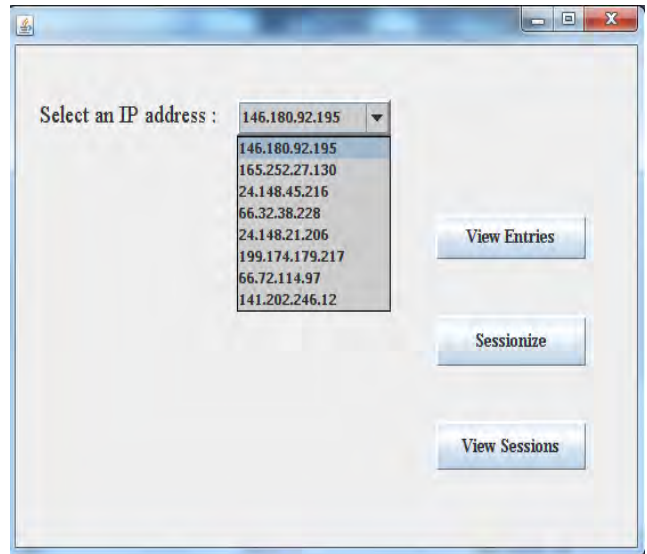


Figure 10 IP addresses having support count greater than or equal to 30.

The entries of the IP address selected by the user can viewed by clicking on the 'view entries' button.

id	date	time	ip	csystem	username	session	sp	sport	csmethod	csystem	csystem	session	csystem	agent	referer
1	2002-04-04	00:20:50	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
2	2002-04-04	00:20:56	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
3	2002-04-04	00:21:11	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
4	2002-04-04	00:21:16	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
5	2002-04-04	00:21:26	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
6	2002-04-04	00:21:30	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
7	2002-04-04	00:22:00	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
8	2002-04-04	00:22:04	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
9	2002-04-04	00:23:13	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
10	2002-04-04	00:23:15	165.252.27.130	Default.asp	-	-	443	116	http://	www.cs.mccall.ca	-	-	-	-	-
11	2002-04-04	00:24:01	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
12	2002-04-04	00:24:15	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
13	2002-04-04	00:24:36	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
14	2002-04-04	00:25:09	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
15	2002-04-04	00:25:22	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
16	2002-04-04	00:25:24	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
17	2002-04-04	00:25:26	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
18	2002-04-04	00:26:32	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
19	2002-04-04	00:26:32	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
20	2002-04-04	00:26:40	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
21	2002-04-04	00:26:41	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
22	2002-04-04	00:27:02	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
23	2002-04-04	00:27:03	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
24	2002-04-04	00:27:04	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
25	2002-04-04	00:27:36	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
26	2002-04-04	00:28:12	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
27	2002-04-04	00:40:59	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
28	2002-04-04	00:41:14	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
29	2002-04-04	00:42:10	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-
30	2002-04-04	00:46:17	165.252.27.130	Default.asp	-	-	80	116	http://	www.cs.mccall.ca	-	-	-	-	-

Figure 11 Entries of IP address selected from the drop down menu.

Sessions are identified. The session time is taken to be 5 minutes.

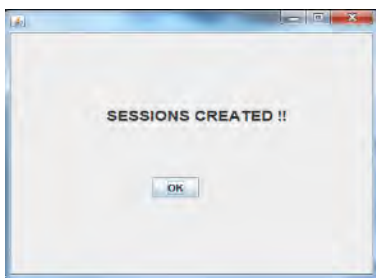


Figure 12 Screen which appears after sessions are created.

Depending on the pages requested the entries of each session are classified among 5 different predefined classes. The entries and the classification of entries of a session are shown in figure below.

ID	Time	IP	Location	Address	Duration
1	00:20:59	165.252.27.130	165.252.27.130	118	
2	00:20:40	165.252.27.130	165.252.27.130	118	
3	00:21:10	165.252.27.130	165.252.27.130	118	
4	00:21:10	165.252.27.130	165.252.27.130	118	
5	00:21:20	165.252.27.130	165.252.27.130	118	
6	00:21:10	165.252.27.130	165.252.27.130	118	
7	00:21:10	165.252.27.130	165.252.27.130	118	
8	00:21:10	165.252.27.130	165.252.27.130	118	
9	00:21:10	165.252.27.130	165.252.27.130	118	
10	00:21:10	165.252.27.130	165.252.27.130	118	
11	00:21:10	165.252.27.130	165.252.27.130	118	
12	00:21:10	165.252.27.130	165.252.27.130	118	
13	00:21:10	165.252.27.130	165.252.27.130	118	
14	00:21:10	165.252.27.130	165.252.27.130	118	
15	00:21:10	165.252.27.130	165.252.27.130	118	
16	00:21:10	165.252.27.130	165.252.27.130	118	
17	00:21:10	165.252.27.130	165.252.27.130	118	
18	00:21:10	165.252.27.130	165.252.27.130	118	
19	00:21:10	165.252.27.130	165.252.27.130	118	
20	00:21:10	165.252.27.130	165.252.27.130	118	
21	00:21:10	165.252.27.130	165.252.27.130	118	
22	00:21:10	165.252.27.130	165.252.27.130	118	
23	00:21:10	165.252.27.130	165.252.27.130	118	
24	00:21:10	165.252.27.130	165.252.27.130	118	
25	00:21:10	165.252.27.130	165.252.27.130	118	
26	00:21:10	165.252.27.130	165.252.27.130	118	
27	00:21:10	165.252.27.130	165.252.27.130	118	
28	00:21:10	165.252.27.130	165.252.27.130	118	
29	00:21:10	165.252.27.130	165.252.27.130	118	
30	00:21:10	165.252.27.130	165.252.27.130	118	
31	00:21:10	165.252.27.130	165.252.27.130	118	

The entries of this session are found to belong to following classes: Class A, Class C, Class B

Figure 13 Entries belonging to session of 5 minutes and the classes to which the entries are classified.

After all sessions are viewed, the association rules can be viewed. These association rules show the relation between the IP address and the pages requested by the clients from that IP address. The association rules for IP address 165.252.27.130 are shown in the following figure.



Figure 14 Association rules for the IP address 165.252.27.130

V. CONCLUSION

Web Usage Mining is an aspect of data mining that has received a lot of attention in recent years.

In this paper, implementation of a system for pattern discovery using association rules is discussed as a method for Web Usage Mining. Different transactions that are closely related to each other are grouped together by the use of clustering approaches on the preprocessed dataset.

The analysis of such clusters will lead to discovery of strong association rules. We obtained all significant association rules between items in the large database of transactions. The relation between different page requests was found.

The support and the confidence values of extracted rules are considered for obtaining the interest of the web visitors. Consequently, the number of hit can be increased by analyzing the visitor attitude.

The approach discussed in this paper, helps the web designers to improve their website usability by determining related link connections in the website.

REFERENCES

- [1] M. Henri Briand, M. Fabrice Guillet, M. Patrick Gallinari, M. Osmar Zaaiane, "Web Usage Mining: Contributions to Intersites Logs Preprocessing and Sequential Pattern Extraction with Low Support", World Academy of Science, Engineering and Technology 48 2008.
- [2] Mr. Sanjay Bapu Thakare, Prof. Sangram. Z. Gawali, "A Effective and Complete Preprocessing for Web Usage Mining", Expert Systems with Applications, 36(3), 6635-6644.
- [3] Resul Daş, Ibrahim Türkoğlu, "Extraction of Interesting Patterns through Association Rule Mining For Improvement of Website Usability", Proceedings of the 2006 IEEE/WIC/ACM International Conference of Web Intelligence (WI 2006 Main Conference Proceedings) (WI'06) 2006 IEEE.
- [4] Bamshad Mobasher, Namit Jain, Eui-Hong (Sam) Han, Jaideep Srivastava, "Web Mining: Pattern Discovery from World Wide Web Transaction", Proc. IEEE International Conference Multimedia Computing Systems, Hiroshima, Japan, June, 1996.
- [5] B. Mobasher, R. Cooley, and J. Srivastava, "Automatic personalization based on Web usage mining" Communications of the ACM, vol. 43, pp. 142-151, 2000.
- [6] C. R. Anderson, P. Domingos, and D. S.Weld, "Adaptive Web Navigation for Wireless Device" Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence, pp. 879-884, 2001.
- [7] I. Cadez, D. Heckerman, C. Meek, P. Smyth, and S. White, "Visualization of navigation patterns on a Web site using model-based clustering," Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, pp.280-284, 2000.
- [8] Dr.R.Lakshminpathy, V.Mohanraj, J.Senthilkumar, Y.Suresh, "Capturing Intuition of Online Users using a Web Usage Mining" Proceedings of 2009 IEEE International Advance Computing Conference (IACC 2009)Patiala, India, 6-7 March 2009.
- [9] Kiruthika M, Dipa Dixit, Pranay Suresh, Rishi M. "An Approach to Convert Unprocessed Weblogs to Database Table"
- [10] "Identifying User Behavior by Analyzing Web Server Access Log File" by K R Suneetha, R Krishnamoorthi.

The macroeconomic effect of the information and communication technology in Hungary

Peter Sasvari

Institute of Business Sciences, Faculty of Economics
University of Miskolc
Miskolc, Hungary

Abstract— It was not until the beginning of the 1990s that the effects of information and communication technology on economic growth as well as on the profitability of enterprises raised the interest of researchers. After giving a general description on the relationship between a more intense use of ICT devices and dynamic economic growth, the author identified and explained those four channels that had a robust influence on economic growth and productivity. When comparing the use of information technology devices in developed as well as in developing countries, the author highlighted the importance of the available additional human capital and the elimination of organizational inflexibilities in the attempt of narrowing the productivity gap between the developed and developing nations. By processing a large quantity of information gained from Hungarian enterprises operating in several economic sectors, the author made an attempt to find a strong correlation between the development level of using ICT devices and profitability together with total factor productivity. Although the impact of using ICT devices cannot be measured unequivocally at the microeconomic level because of certain statistical and methodological imperfections, by applying such analytical methods as cluster analysis and correlation and regression calculation, the author managed to prove that both the correlation coefficient and the gradient of the regression trend line showed a positive relationship between the extensive use of information and communication technology and the profitability of enterprises.

Keywords: *ICT; Economic sector; Profitability; Total Factor Productivity.*

I. INTRODUCTION

Despite its fast expansion and more widespread usage, the effect of information and communication technology on productivity was negligible in the 1980s as well as in the beginning of the 1990s [1]. Later, more and more studies were published arguing that information and communication technology contribute to the growth of productivity.

The effects of ICT on economic growth got into the centre of interest due to the technological revolution taking place in the US in the second half of the 1990s. The previous international productivity trends were altered in the period 1996-2000. The pace of productivity growth, which is defined by production per worker or rather production per labour hour,

accelerated in the US and, in contrast to the previous decades, exceeded the growth dynamism of the European Union including its economic and monetary union. With the bursting of the dotcom bubble (the dramatic decrease in the value of blue chip stocks), the analysis of the macro- and microeconomic effects of ICT was overshadowed. Part of the reason of this moderate attention was that considerable excess capacity had become available in the blue chip sector and the real economy needed some time to make use of it. In the period of the recession, or rather of the low growth of GDP, following the bursting of the technological bubble, companies were forced to cut back on their investments in information and communication technology. The acceleration of the growth of productivity made the pace of GDP growth more dynamic as well.

According to the European Committee's report, the macroeconomic effects of information and communication technology predominate in four channels, having an influence on economic growth and productivity [7].

1. The production of information and communication technologies is inseparable from rapid technological development,
2. The investment channel, that is, the growth of production potential by accumulating information and communication technology capital,
3. The possible production externalities (network externalities, network of external economic effects) connected to information and communication technology,
4. The increased demand for information and communication technology may intensify the demand for other types of labour force and capital. However, it must be taken into consideration that ICT is a substitution for other outputs, and the application of new technologies with the necessary restructurization leads to tensions in the labour and capital markets.

This is the reason why negative growth effects should be taken into account in the case of the fourth channel, at least in the short and medium term. In many cases, information technology may soften the demand for human resources because ICT products themselves can substitute the living work or embodied human capital used to produce them. The

The described work was carried out as part of the TÁMOP-4.2.1.B-10/2/KONV-2010-0001 project in the framework of the New Hungarian Development Plan. The realization of this project is supported by the European Union, co-financed by the European Social Fund.

substitution of human resources with information and communication technologies will probably increase in the future as technologies and delivery systems become more mature.

In theory, the owners of information and communication technologies can benefit from the information revolution by gaining higher profits, employees can achieve higher salaries or wages, and users can also feel the beneficial effects in the form of lower prices. Based on empirical experiences, profits and salaries showed a modest increase but these changes were slight in comparison to the drop in the relative price of ICT products. This allows us to conclude that the countries employing information technologies benefited slightly more from the information technology revolution than ICT-producing countries as part of their profits were lost due to the unfavourable changes in purchasing power parity. It must be noted that salaries and wages paid in the ICT sector are higher in a great number of developing countries manufacturing information technologies than elsewhere [3].

For the time being, the effect of information technologies is not unambiguous on the geographical location of production in the relation between centre and periphery. On the one hand, as transaction and communication costs become lower, the flexibility of production comes to the front against economy of scale opportunities, a robust increase in the expansion of economic activities; a kind of deconcentration can be expected. On the other hand, the more precise information available on the changes of consumer preferences, the growing weight of intermediate goods used as inputs in the production of other products, and the opportunity of outsourcing certain economic activities to suppliers all make it advantageous for companies to produce close to markets. If outsourcing results in a growing number of new intermediaries providing various partly customized services (accounting, marketing, purchasing etc.), being close to markets is also essential as concerned companies are able to save up more time [4].

The use of information technology products continues to expand rapidly in the developed countries but productivity advantages appear more slowly than in the case of developing countries. Similar to developed nations, the reducing price of information technology products is likely to be the main driving force in the developing world. The use of information technology can be detected in the growth of productivity only if the additional human capital is available, the deregulation of telecommunication infrastructure and information flow takes place and it is possible to eliminate organizational inflexibilities that prevent companies from exploiting the advantages of new technologies and ideas. Although information technologies contribute to raising the standard of productivity in the developing world, the productivity gap may widen between the developed and developing countries.

The high standard of human capital shows a strong correlation with the adaptation of information technologies. As new technologies usually appear in the form of new equipment, high investment rates accelerate their adaptation.

Finally, a strong connection could be observed between the expansion of information technologies and economic policy. The probability of the expansion of new technologies is higher

when economic policies are open to imports and the inflow of foreign working capital.

The rate of Internet users and the number of mobile phones is not necessarily lower in some of the poorest countries than in the countries representing a higher economic development level. This allows us to conclude that there is a strong intention to have access to information technologies and international knowledge networks even in the poorer countries. The real question is whether these technologies can be used for accelerating economic growth in those nations.

Information technologies provide attractive opportunities to "by-pass" and exceed out-of-date technologies.

Modern technologies also bring education much faster to a considerably larger number of people than before. With having better access to information and lower transaction costs, people living in the periphery of domestic and international markets can join the mainstream by using up-to-date information technologies. The opportunities of applying information technologies are outstanding in raising productivity, including plants, banks, ports and even governments. These trends are strengthened by continuous innovations and cost reductions.

The effect of information technologies on productivity is not perceptible in the whole developing world. In many cases, the main reasons for this are the lack of adequate complementary human capital, the low responsiveness of the telecommunication sector and a fair amount of inflexibility. Regarding human capital, information technologies may moderate the demand for human resources as IT products themselves substitute the living labour needed to manufacture them. At the same time, the need for complementary human capital in information technologies can be significant, especially in the field of business and government applications.

II. THE METHOD OF THE RESEARCH

The examination of the subject is interdisciplinary as it has social and scientific references, so a complex approach was needed when I started processing the literature. I needed to study literature on economics, law, sociology and technology connected to the information society.

In consideration of the complexity of the studied subject, I selected several analytical methods and approaches. During the data collection, I reclined upon the Hungarian and the international literature on the subject, thus I was able to process a large quantity of information (nearly 6000 figures). I also extended my literature research to printed and electronic publications on the Internet. As part of my research, I conducted an empirical survey among Hungarian companies and enterprises. The questionnaire was mainly answered by senior directors of the related companies (executive directors, Human Resources managers etc.), in the case of sole proprietorship, sole proprietors themselves as self-employed persons gave the answers to the questions. The questionnaire was filled in by 536 respondents altogether. The sampling unit consisted of Hungarian enterprises operating in several economic sectors; the chosen sampling method was accidental sampling. The applied methods used for processing the primary data of the research were a correlation and regression

calculation, multiple regression models and a customized indicator system in SPSS 16.0 [9].

III. THE MACROECONOMIC EFFECT OF THE INFORMATION AND COMMUNICATION TECHNOLOGY

ICT devices contribute to the improvement of productivity, the economic growth or the acceleration of the economy in several areas. As far as macroeconomic effects are concerned, the technological development is very rapid alongside with the productivity and the total factor productivity (TFP) in the economic sectors producing ICT devices. On the one hand, this process increases the national average in itself, especially when its share tends to grow in the GDP; on the other hand it makes other economic sectors more dynamic by the technological and economic links throughout the whole economic system.

Profits gained with the help of the rapid technological development and the improvement of productivity was eroded by the dropping ICT prices. Countries producing ICT devices lost a part of their profits realized from production because of the deteriorating swap ratio.

The source of productivity and growth benefits from capital deepening (it describes an economy where the amount of capital per worker is increasing), that is the growing rate of using ICT devices, which is stimulated by the huge decrease in ICT prices. These benefits appear in the form of the increased output of existing products and services, manufacturing new products or providing new services, fulfilling customer needs more efficiently and decreasing transition costs etc. As the effect of ICT devices on increased productivity and more dynamic growth are connected to capital deepening, it can be seen that the countries and businesses using these new technologies have benefited more from the revolution of information technology, than the countries producing them.

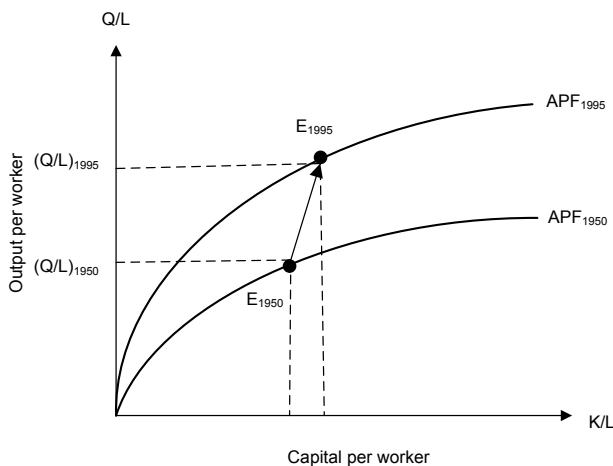


Figure 1. Technological development elongates the production curve [6]

ICT devices also increase the total factor productivity that is they improve the degree of utilization of capital and labour force. The total factor productivity (TFP) is applied to express the overall effect on the savings of economies of scale, management skills, production externalities and other, non-traditional factors influencing productivity. The significance of the growing total factor productivity is that it accelerates the pace of economic growth without any additional costs as well

as without having to increase the quantity input. Capital deepening is a necessary but not sufficient condition for improving productivity. It can only unfold in its fullest form when the potential efficiency surplus of ICT devices is exploited. A more dynamic TFP automatically accelerates the pace of labour productivity, thus it helps to boost economic performance.

Using ICT devices also improves productivity and makes economic growth more dynamic because information technology cannot be regarded as capital goods in the traditional sense of the word. The installation of a new information technology device raises the value of other existing devices as well. As a consequence, network effects may occur within companies, moreover they may appear between industrial sectors, and they may necessitate shaping new forms of cooperation (outsourcing).

As it was stated above, ICT devices increase productivity and output by capital deepening, improved total factor productivity and network externalities at the microeconomic level. The advantages of using ICT devices at the macroeconomic level come from all the advantages of the companies' improved productivity and from the network advantages based on the feature of reducing transition costs and accelerating innovation. The network advantage does not depend on the operation of a given company and its business strategy.

However, the effects of ICT devices on the productivity of companies cannot be measured unequivocally at the microeconomic level because of certain statistical and methodological imperfections, the difficulties in measuring network effect at a business level and the lack of data enabling to make international comparisons. Furthermore, the effects of ICT devices on productivity appear at a later time, as they are preceded by a longer or shorter learning process. The productivity paradox has started to vanish by now. It has become clear that statistics cannot or just partially show the secondary effects of using ICT devices in the economy (faster information processing, improvement of productivity in producing knowledge, for instance).

In countries where competition is fierce in the market, enterprises using ICT devices are not necessarily the main winners of capital deepening, it is the customers who can benefit from it by getting lower prices, better quality or more convenience.

It is not necessarily true in countries where competition is weak. Here, companies are able to realize a greater part of benefits coming from capital deepening. But it has its own price as the secondary effects of using ICT devices are more limited in the economy.

With the help of the compound indicator and the financial data of the studied economic sectors, an attempt was made in the research to find a connection between the development levels of ICT and their profitability. Profitability and productivity are influenced by a lot of other factors as well. As it was not possible to measure and show the effect of those other factors, the results are not full but informative.

Based on the statistical connection between the compound indicator and the increment of the Gross Value Added per worker, the correlation coefficient is 0.13, while the gradient of the regression trend line is 0.17. Both numbers show a positive connection between the compound indicator and profitability [8].

Then, using a coordinate system, the connection between the changes of the specific indicators of the studied economic sectors and the development level of those sectors was illustrated. The Y axis shows the growth pace of Gross Value Added per capita in the economic activities between 2003 and 2008 [10]. The X axis shows the compound indicator that was created for measurement purposes. The points defined by the two values show clearly where a given economic sector can be found in the coordinate system, what groups can be constituted, and what tendency can be observed.

The highest increment of specific Gross Value Added was produced by the sectors ‘Manufacturing’, ‘Electricity, gas and water supply’, ‘Transport, storage and communication’ and ‘Financial intermediation’. With the exception of ‘Electricity, gas and water supply’, all of these economic activities belong to the group of underdeveloped sectors (below 50%).

High (but still not reaching the developed status) compound indicators were shown by the sectors ‘Mining and quarrying’ and ‘Wholesale and retail trade; repair work’, as they produced an increment of Gross Value Added below the average, these economic sectors can be found in the lower right part of the coordinate system. The sectors ‘Construction’, ‘Health and social work’ and ‘Hotels and restaurants’ can be seen as laggards, so they got into the lower left part of the coordinate system.

The ‘Agriculture, hunting and forestry’ sector can also be classified as a laggard economic activity, but as the effect of the compound indicator on the increment of Gross Value Added was less significant, it can be found in the upper left part of the coordinate system. Drawing a trend line on the points, it is clear that the line shows a positive gradient, that is, the higher the usage of ICT devices, the higher improvement can be detected in the specific Gross Value Added.

IV. DETERMINING THE NET INCOME OF ENTERPRISES BY USING MULTIPLE LINEAR REGRESSION INCLUDING THE OTHER VARIABLES

The connection between several socioeconomic phenomena is so complicated that the change of a given variable cannot be characterized sufficiently in each case with the help of another variable related to the former one. The analysis has to be extended to many criteria even when the aim of the analysis is only to understand the connection between two criteria, as it is very rare in the economy when the link between two phenomena can be studied separate from other essential effects [5].

The specific methods of studying multiple connections are correlation analysis and multiple regression analysis. The former method measures the strength of the arithmetic relationship between two variables, while the aim of the latter method is to find a standard pattern in stochastic relationships.

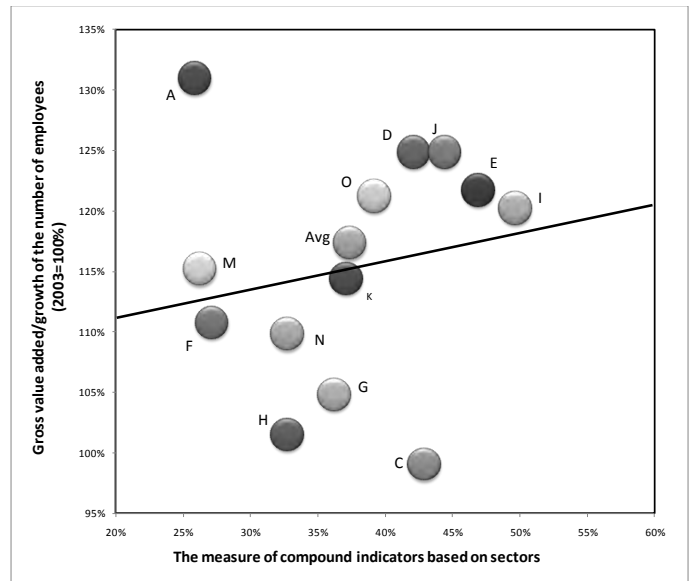


Figure 2. Connection between the growth of gross value added and the development level of information and communication technology in several economic activities¹

Correlation analysis describes the strength and direction of a linear relationship between two or more variables. In the correlation analysis presented below, I used metric variables.

In my primary research, I expressed the variable of the net income of enterprises with the linear combination of the following continuous variables:

headcount data that are further divided into the number of IT specialists employed, the total number of employees, the number of regular personal computer users, the number of employees using computers connected to the Internet; and the e-commerce financial data such as the value of purchases via the Internet, the sales revenues via the Internet, the value of purchases through computer networks and the sales revenues through computer networks expressed in thousand forints.

Independent variables	Regression analysis	Dependent variable
1. Headcount data: 1.1. The number of IT specialists employed 1.2. The total number of employed 1.3. The number of regular personal computer users 1.4. The number of employees using computers connected to the Internet 2. E-commerce financial data: 2.1. The value of purchases via the Internet 2.2. Sales revenues via the Internet 2.3. The value of purchases through computer networks 2.4. Sales revenues through computer networks	Multiple linear regression	Net sales income of enterprises

Figure 3. Net income of enterprises expressed with multiple linear regression [5]

The table below shows that seven independent variables got into the multiple regression models with the exception of the sales revenues via the Internet.

¹ A=Agriculture, hunting and forestry, C=Mining and quarrying, D=Manufacturing, E=Electricity, gas and water supply, F=Construction, G=Wholesale and retail trade; repair work, H=Hotels and restaurants, I=Transport, storage and communication, J=Financial intermediation, K=Real estate, renting and business activities, M=Education, N=Health and social work, O=Other community, social and personal service activities.

TABLE I. IDENTIFICATION OF VARIABLES ENTERED OR REMOVED INVOLVED IN THE RESEARCH

Model	Variables Entered	Variables Removed	Method
1	1.1. The number of IT specialists employed 1.2. The total number of employed 1.3. The number of regular personal computer users 1.4. The number of employees using computers connected to the Internet 2.1. The value of purchases via the Internet 2.3. The value of purchases through computer networks 2.4. Sales revenues through computer networks		Enter

Performing a multiple linear regression analysis makes sense only when the stochastic relationship can be demonstrated between the independent variables and the dependent variable. It can be seen in the table that the resulting multiple linear regression model is good because the value of R2 statistics is 0.999 which means that the given model explains 99.9% of all variances.

TABLE II. SUMMARY OF MULTIPLE LINEAR REGRESSION ANALYSIS (MODEL SUMMARY)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	1,000	,999	,998	34069,514

The analysis of variance shows the significance of the part explained by the applied linear model. In other words it shows whether a model has an explanatory power or not. The second column represents the sum of squares. Total row of the table presents the total sum of squares. The Residual row of the table refers to the partial sum of squares. The third column shows the degrees of freedom. The F-test in the analysis of variance is used to assess whether the expected values of a quantitative variable within several pre-defined groups differ from each other. It serves as a measure of the statistical importance or significance of the differences among the group means. Mean square is the quotient of the sum of squares and degrees of freedom, the values of mean squares are shown in the fourth column. The F-test can be found in the next column, its value determines the value of significance shown in the sixth column. If the value of significance is low (less than 0.005), I have to leave the null hypothesis according to which the independent variables and the dependent variable have no relationship with one another.

It can be proven that the independent variable explains a significant part of the dependent variable, so it is worth studying the strength of their relationship, in other words, the size of the explained part.

As the null hypothesis proved to be true, the result of the following t-test shows that the variable "the number of IT specialists employed" has no significance in the model because its significance level is high.

TABLE III. REGRESSION COEFFICIENTS IN THE CASE OF USING ENTER METHOD (COEFFICIENTS)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-74879,4	15734,4		-4,7	,018
	1.3. The number of regular personal computer users	76222,8	6852,5	4,6	11,1	,002
	2.1. The value of purchases via the Internet	1,108	,120	,363	9,2	,003
	2.3. The value of purchases through computer networks	43,73	6,157	1,813	7,1	,006
	2.4. Sales revenues through computer networks	-111,13	8,872	-2,973	-12,5	,001
	1.1. The number of IT specialists employed	88415,08	67336,68	,402	1,3	,281
	1.2. The total number of employed	13672,86	2183,621	1,514	6,3	,008
	1. 4. The number of employees using computers connected to the Internet	-67251,95	4075,924	-4,124	-16,5	,000

Using the backward method, the removal of the independent variables from the model continues until the partial explanation of all remaining variables has significance.

At first, every independent variable appears in the model, however, it can be seen that one of them does not have a significance in the explanation of the net income of enterprises. In the first step, the variable „the number of IT specialists employed" was left out of the model as it produced the smallest t-value in absolute terms. The six remaining variables in the final model proved to be significant.

The next step is the interpretation of Beta values. Its constant value is -79497. This value denotes where the axis of the net income of enterprises is cut by the hyperplane containing the six remaining variables.

This hyperplane cuts the axis of the main income only if the value of the six remaining variables is 0. The strongest effect may be caused by the increase of the number of regular personal computer users on enterprises. If the value of the purchases on the Internet and through computer networks increases by 1000 forints, the net income of enterprises will rise by 51000 forints (2.1. and 2.3.).

Another interesting feature is that the variable 'sales revenues through computer networks' has a negative effect on the net income of enterprises.

TABLE IV. REGRESSION COEFFICIENTS IN THE CASE OF BACKWARD METHOD (COEFFICIENTS)

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error			
1 (Constant)	-74879,4	15734,5		-4,76	,018
1.3. The number of regular personal computer users	76222,8	6852,6	4,65	11,12	,002
2.1. The value of purchases via the Internet	1,1	,12	,36	9,23	,003
2.3. The value of purchases through computer networks	43,7	6,2	1,81	7,10	,006
2.4. Sales revenues through computer networks	-111,1	8,8	-2,97	-12,53	,001
1.1. The number of IT specialists employed	88415,1	67336,6	,402	1,31	,281
1.2. The total number of employed	13672,9	2183,6	1,51	6,26	,008
1. 4. The number of employees using computers connected to the Internet	-67251,9	4075,9	-4,12	-16,50	,000
2 (Constant)	-79497,0	16666,8		-4,77	,009
1.3. The number of regular personal computer users	73207,6	7016,4	4,46	10,43	,000
2.1. The value of purchases via the Internet	1,3	,07	,41	17,19	,000
2.3. The value of purchases through computer networks	50,2	4,03	2,08	12,45	,000
2.4. Sales revenues through computer networks	-106,5	8,84	-2,85	-12,05	,000
1.2. The total number of employed	16249,9	1040,1	1,79	15,62	,000
1. 4. The number of employees using computers connected to the Internet	-67639,1	4417,9	-4,14	-15,31	,000

The calculation of beta values all variables are entered into the model in a standardized form with 0 mean and with unit square. Beta values indicate which independent variable has a stronger effect or higher significance, in the case of my model the beta value of the variable 'the number of regular personal computer users' has the strongest effect (4460). It also turned out that the variables 'the number of employees using computers connected to the Internet' and 'sales revenues through computer networks' did not increase the net income of enterprises.

The table below contains the variables excluded from the model.

TABLE V. TABLE 1 - EXCLUDED VARIABLES

Model	Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance
1					
2.2. Sales revenues via the Internet	,053	,024	,983	,017	6,621E-5
2					
2.2. Sales revenues via the Internet	,918	1,149	,334	,553	,000
1.1. The number of IT specialists employed	,402	1,313	,281	,604	,002

V. CONCLUSIONS AND SUGGESTIONS FOR THE PRACTICAL USE OF RESEARCH FINDINGS

It was shown by the figures that one possible reason for gaining advantages from using ICT devices at the macroeconomic level come from the improved productivity of enterprises and from the network advantages based on the feature of reducing transition costs and accelerating innovation. As a consequence, ICT devices increase productivity and output by capital deepening, improved total factor productivity and network externalities at the microeconomic level. At the same time, it has become clear that statistics cannot or just partially show the secondary effects of using ICT devices in the economy as a whole (for instance, faster information processing, improvement of productivity in producing knowledge.)

However, I was able to show the positive correlation between the use of ICT devices and the profitability of the Hungarian enterprises in the studied economic sectors as both the correlation coefficient and the gradient of the regression trend line referred to a strong connection between the compound indicator and the profitability of enterprises. Although the end conclusion of the research is that the use of information and communication technology in the studied Hungarian economic sectors can still be regarded as underdeveloped, with the exception of the agricultural sector where the effect of the compound indicator was the least significant, the positive gradient of the trend line showed that the use of ICT devices resulted in a higher improvement in the specific Gross Value Added in almost every Hungarian economic sector.

As it can be predicted that information and communication technology will have an even deeper impact on the operation and profitability of enterprises in the near future, further researches should be conducted in order to measure its scope and extent.

REFERENCES

- [1] M. Castells, "The Information Age," Gondolat- Infonia, 2000.
- [2] A. Kápolnai, A. Nemeslaki, and R. Pataki, "eBusiness stratégia vállalati felsővezetőknek (E-business strategies for senior management)," Aula, 2002.
- [3] Z. L. Karvalics, "Információ, társadalom, történelem, Válogatott írások, (Information, society, history, Selected works)," Typotex Kiadó, 2003.

- [4] Z. L. Karvalics, "Információs társadalom – mi az? Egy kifejezés jelentése, története és fogalomkörnyezete (Information Society – what is it exactly? The meaning, history and conceptual framework of an expression), Az információs társadalom, Az elmélettől a politikai gyakorlatig (The Information Society, From theory to political practice)," Gondolat – Új Mandátum, Budapest, 2007, pp. 29-47
- [5] L. Sajtos, A. Mitev, "SPSS kutatási és adatelemzési kézikönyv (The handbook of SPSS research and data analysis)," Alinea Kiadó, Budapest, 2007.
- [6] P. A. Samuelson, W. D. Nordhaus, "Közgazdaságtan (Economics)," Akadémiai Kiadó, 1998.
- [7] K. Szabó, B. Hámori, "Információgazdaság (Information Economy)," Akadémiai Kiadó, Budapest, 2006.
- [8] P. Sasvari, "A Comprehensive Survey on the Economic Effects of Information and Communication Technology in Hungary Social Sciences Bulletin", Sociālo Zinātņu Vēstnesis 2011 1, 7-25 p.
- [9] P. Sasvari, "The State of Information and Communication Technology in Hungary – A Comparative Analysis", Informatica 35, 2011, Slovenia, 239-244 p.
- [10] Hungarian Central Statistical Office, Available: <http://www.ksh.hu>

AUTHOR PROFILE



Dr. Peter Sasvari received his MSc in Mechanical Engineering, MSc in Economics and his PhD in Business and Organisation Sciences at the University of Miskolc. Now he is an associate professor at the Institute of Business Science, Faculty of Economics, University of Miskolc, Hungary. His current research interests include different aspects of Economics, ICT and the information society.

Preprocessor Agent Approach to Knowledge Discovery Using Zero-R Algorithm

Inamdar S. A

School of Computational Science
Swami Ramanand Teerth,
Marathwada University, Nanded

Narangale S.M.

School of Media Studies
Swami Ramanand Teerth,
Marathwada University, Nanded

G. N. Shinde*

Indira Gandhi College
CIDCO, Nanded-431603,
Maharashtra, India

Abstract— Data mining and multiagent approach has been used successfully in the development of large complex systems. Agents are used to perform some action or activity on behalf of a user of a computer system. The study proposes an agent based algorithm PrePZero-r using Zero-R algorithm in Weka. Algorithms are powerful technique for solution of various combinatorial or optimization problems. Zero-R is a simple and trivial classifier, but it gives a lower bound on the performance of a given dataset which should be significantly improved by more complex classifiers. The Proposed Algorithm called PrePZero-r has significantly reduced time taken to build the model than Zero-R algorithm by removing the Lower Bound Values 0 while preprocessing and comparing the result with class values. Also proposed study introduced new factor “Accuracy (1-e)” for each individual attribute.

Keywords- Data mining; Zero-R algorithm; Lower Bound Value; Class values.

I. INTRODUCTION

Data mining has various techniques to extract useful information in large amounts of data. Data mining is defined as a technique of finding hidden information in a database [1]. It may be called as data driven discovery, explorative data analysis, deductive learning. Data mining in general falls in to the following categories: classification patterns, association patterns, sequential patterns, and spatial-temporal patterns. The important feature of Data Mining algorithms is running time of an algorithm must be predictable and acceptable in large database.

II. RELATED WORK

Research has shown that over the past few year data mining tools are heavily used in healthcare spectrum. Agent based approach has become an advanced trend in Knowledge discovery. The Classify Agent offers an alternative to achieving the data mining purpose of obtaining a good model with faster classification time from large database within reasonable timeframe. In Agent Based Meta Model Agents are basic modeling entities that maintain a set of properties and behaviors. By factoring agents, relationships, and behaviors into separate components, more modular and expressive models can be created. Research shows the knowledge discovery is using multi-agent approach for quicker and reliable information retrieval [2-11].

III. PRELIMINARIES

KDD (Knowledge Discovery from Databases) is the process of finding useful information and patterns in data. Data mining is the use of algorithms to extract the information and patterns derived by the KDD process. KDD is multistep process, the input to this process is the data and the output is the useful information desired by the users [1].

There are many techniques for classification such as neural networks, Bayesian, decision tree, instance based learning, genetic algorithm, rough set, and fuzzy logic [3].

A. Agents:

Agents are used to perform some action or activity on behalf of a user of a computer system. Agent refers to the entities which run in dynamic environment and have higher self-government capacity. Agent software is a type of computer program which simulates human intelligence behavior. Agent should be able to learn from experience and to act autonomously to the ever changing task.

B. MAS:

A group of agents can collectively and collaboratively form a Multi Agent System (MAS) to perform complex and lengthy tasks [1-7].

IV. WHY ZERO-R?

Rules can be extracted from the tree by search the tree path from root to the leaf. C4.5 is an algorithm used to generate a decision tree and an extension of Quinlan's earlier ID3 algorithm [8-12]. C4.5 is a technique to generalize rules associated with a tree that accumulate all the tests between the root node and the leaf node. This technique uses the training dataset to estimate the accuracy of each rule [2, 13-14]. The decision-formation and tree method like the nearest neighbors method, exploits clustering regularities for construction of decision-tree representation.

The decision tree learning method requires the data to be expressed in the form of classified examples. Genetic Algorithms are powerful technique for solution of various combinatorial or optimization problems. They are more an instrument for scientific research rather than a tool for generic practical data analysis. Rough classifier is an extension of logic and discrete mathematics from rough set theory [7]. Like decision tree, rough classifier is a nonparametric model which

suits for the exploratory knowledge discovery and without intervention from users.

In WEKA Zero-R is a simple classifier. Zero-R is a trivial classifier, but it gives a lower bound on the performance of a given dataset which should be significantly improved by more complex classifiers. As such it is a reasonable test on how well the class can be predicted without considering the other attributes. It can be used as a Lower Bound on Performance. Any learning algorithm in WEKA is derived from the abstract WEKA classifiers.

V. THE PROPOSED ZERO-R ALGORITHM

The Proposed Algorithm called PrePZero-R shows that we are removing lower bound values 0 and checking the results how it affects the class value. Following are the steps of PrePZero-R Algorithm shown in fig.1

In figure.1, in first step, we are selecting classifier Zero-R, simultaneously checking for its capabilities, if it does not satisfy its condition capabilities then directly Exit. If it is capable then build classifier. In next step getting Instance values, then set the Lower Boundary value for instance, remove Lower Boundary value form calculation, finally calculate Zero-R class value then Exit.

VI. EXPERIMENTAL SETUP

The experiment was conducted using the UCI Pima Indian data set [2]. We used Data mining library WEKA 3.6.5. In WEKA we introduce new term Accuracy (1-e) which gives better results in error.

VII. RESULT & DISCUSSION

The experiment was conducted using the UCI Pima Indian data set [5]. The dataset contains 768 instances of Pima Indian heritage females who were diagnosed for diabetes. The diagnostic result (diabetes negative or diabetes positive) in the data set. The five attributes are as follow: number of times pregnant, plasma glucose concentration, serum insulin, diabetes pedigree function and finally the test result.

The experiment is to measure time and accuracy of a classification on the UCI Pima Indian dataset [6]. Class variable value is mutually exclusive, either diabetes negative or diabetes positive. There are 4 standard methods for Data Mining: association, classification, clustering techniques and prediction.

In Table 1 the class value for each attribute is compared in Zero-R and PrePZero-R algorithm. The Accuracy is measured in Time (Nano-Seconds) in comparison with Zero-R and PrePZero-R algorithm and the difference is shown in the last column in Table 1.

For most medical applications the logical rules are not precise but vague and the uncertainty is present both in premise and in the decision. For this kind of application a good methodology is the rule representation from decision-tree method, which is easily understood by the user [4].

The experimental result shows that we are removing lower bound values 0 and checking the results how it affects the class value. As shown in table1 the proposed algorithm has

significantly reduced the running time and Accuracy. This criterion is important in agent based data mining to obtain the good knowledge model from the complex and large database. This classifier simply predicts the majority class in the training data. it makes little sense to use this scheme for prediction, it can be useful for determining a baseline performance as a benchmark for other learning schemes. Zero-R tests how well the class can be predicted without considering other attributes. It can be used as a Lower Bound on Performance.

VIII. FUTURE SCOPE

Execution of this PrePZero-R algorithm experiment on parallel multiprocessor system will increase efficiency. Creation and implementation of fuzzy set algorithms tends to increase the accuracy of the output. This experiment assumes the dataset can be minimized for lower bound values while calculating class value.

If fuzzy min-max algorithm is implemented for removal of attribute values from testing parameters, the accuracy can be increased and the time taken to build the model will definitely be reduced. Thus preprocessor agent approach shall be used effectively.

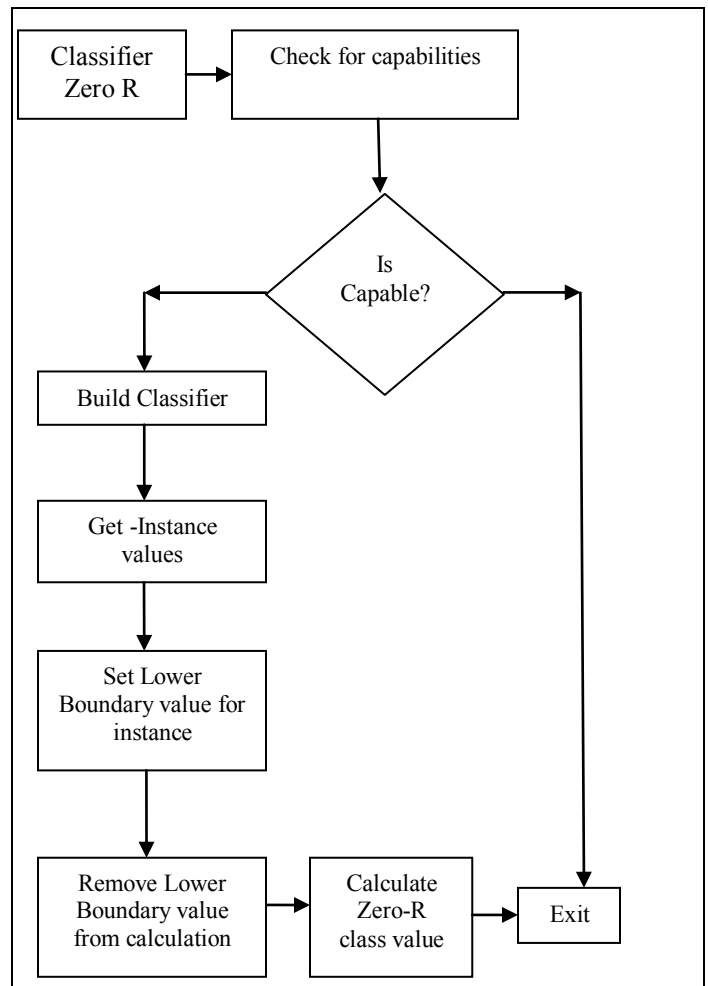


Figure 1. Preprocessing on individual attribute (Removal of lower boundary value from calculation)

TABLE I: COMPARISON OF RESULTS

Attribute	Class Value		Accuracy		Time (Nano Sec)		Difference
	ZeroR	PrePZeroR	ZeroR	PreP ZeroR	ZeroR	PreP ZeroR	
Number of times pregnant	3.8450520833333335	4.494672754946728	97.2255	97.0705	3445410	466400	2979010
Plasma glucose concentration	120.8945313	120.5388128	74.7986	74.8313	578705	462559	116146
Serum insulin	79.79947917	79.48249619	15.3368	15.4382	403892	437346	-33454
Diabetes pedigree function	0.471876302	0.463604262	99.7524	99.7542	523181	17540777	17017596
Test result	0.348958333	0.350076104	99.5452	99.5448	421213	3184273	-2763060

IX. CONCLUSION

To conclude the preprocessing approach for knowledge discovery shows significant increase in performance. The proposed PrePZero-R algorithm using WEKA has significantly reduced running time and increases accuracy. This algorithm does so by removing the Lower Bound Values and comparing the result with class values for each individual attribute. This criterion is important in Agent Based data mining to obtain the good knowledge from complex and large databases.

REFERENCES

[1] Margaret H. Dunham and Sridhar, Data Mining, Introduction and Advanced Topics, Prentice Hall Publication, ISBN 81-7758-785-4
[2] M. Wooldridge, An Introduction to MultiAgent Systems. John Wiley & Sons Ltd, 2002. John Wiley & Sons, 2002.
[3] Abad-Grau, M. M., Arias-Aranda, D.: Operations Strategy and Flexibility: modeling with Bayesian Classifier. 106, 460--484 (2006).
[4] Plamena Andreeva, Maya Dimitrova, Petia Radeva, DATA MINING LEARNING MODELS AND ALGORITHMS FOR MEDICAL

APPLICATIONS

[5] IDI. International diabetes institute - diabetes research, education and care. 2007(10/30/2007).
[6] U. P. Indian, "Pima Indians Diabetes Data Set." vol.2008,2008.
[7] Lenarcik, A., Piasta, Z.: Rough Classifier. In: Ziarko, W. (eds.) Rough Sets, Fuzzy Sets and Knowledge Discovery, pp. 298--316. Springer, London (1994).
[8] Ruggieri, S.: Efficient C4.5. In: IEEE Transactions on Knowledge and Data Engineering, pp. 438--444. IEEE Educational Activities Department, USA (2002).
[9] Han, J., Kamber, M.: Data Mining: Concepts and Techniques. Morgan Kaufmann, San Francisco (2001).
[10] Azuraliza Abu Bakar, Zulaiha Ali Othman, Abdul Razak Hamdan, Rozianiwati Yusof, Ruhaizan Ismail, 'Agent Based Data Classification Approach for Data Mining', 2008 IEEE
[11] Cuong Tong, Dharmendra Sharma and Fariba Shadabi, 'A Multi-Agents Approach to Knowledge Discovery', 2008 IEEE.
[12] Li Zhan, Liu Zhijing, 'Web Mining Based On Multi-Agents', COMPUTER SOCIETY, IEEE(2003)
[13] M. Wooldridge, An Introduction to MultiAgent Systems. John Wiley & Sons Ltd, 2002. John Wiley & Sons, 2002.
[14] IDI. International diabetes institute - diabetes research, education and care.

Text Independent Speaker Identification using Integrated Independent Component Analysis with Generalized Gaussian Mixture Model

N M Ramaligeswararao
Department of Electronics &
Communication Engineering,
GIET affiliated to JNTUK,
Rajahmundry, India

Dr.V Sailaja
Department of Electronics &
Communication Engineering,
GIET affiliated to JNTUK,
Rajahmundry, India

Dr.K. Srinivasa Rao
Department of statistics,
Andhra University
Visakhapatnam, India

Abstract - Recently much work has been reported in literature regarding Text Independent speaker identification models. Sailaja et al (2010)[34] has developed a Text Independent speaker identification model assuming that the speech spectra of each individual speaker can be modeled by Mel frequency cepstral coefficient and Generalized Gaussian mixture model. The limitation of this model is the feature vectors (Mel frequency cepstral coefficients) are high in dimension and assumed to be independent. But feature represented by MFCC's are dependent and chopping some of the MFCC's will bring falsification in the model. Hence, in this paper a new and novel Text Independent speaker identification model is developed by integrating MFCC's with Independent component analysis(ICA) for obtaining independency and to achieve low dimensionality in feature vector extraction. Assuming that the new feature vectors follows a Generalized Gaussian Mixture Model (GGMM), the model parameters are estimated by using EM algorithm. A Bayesian classifier is used to identify each speaker. The experimental result with 50 speaker's data base reveals that the proposed procedure outperforms the existing methods.

Keywords - Independent component analysis; Generalized Gaussian Mixture Model; Mel frequency cepstral coefficients; Bayesian classifier; EM algorithm.

I. INTRODUCTION

The growing need for automation in complex work environments and the increased need for voice operated services in many commercial areas have motivated the present work. While many existing systems for speaker identification achieve good performance in relatively constrained environments, performance invariably deteriorates in noisier environment. Speaker identification system is the process of selecting the best matched speaker among the enrolled speakers, with features extracted from speech signals.

Many techniques involving statistical or probabilistic approaches have been applied to speaker specific speech patterns (Leena Mary and Yegnanarayana (2008), Jyoti et al (2011)) [22] [18]. Several methods were employed to separate mixed signals known as 'Blind Source Signals' (BSS) [13]. The term blind refers to the fact that the method of

combination and source signal characteristics are unknown, so BSS permits a wide range of signals as input.

Text independent speaker identification system has many potential applications like security control, telephone banking, information retrieval systems, speech and gender recognition systems, etc. Speaker identification system involves two parts: front-end (feature extractions) and back-end (actual recognition). These system use processed form of speech signals instead of using raw speech signals as it is obtained. This is to reduce the time consumed in identifying the speaker and to make the process easy, by reducing the data stream and exploiting its advantage of being redundant. Computation of cepstral coefficients using preprocessing and feature extraction phases plays a major role in text independent speaker identification systems Ning Wang et al (2010) [24].

Various studies made by Zhu, (1994) and Furui, (1982)[38][28] have also shown that computing cepstral coefficient is the best among all the parameters for any type of speaker recognition. It was proved that the performance of the speech recognizers can improve using cepstral representation of the signals for both clean and noisy speech (Erell, (1993)) [11]. Recently Reynold (1994)[5] have used Mel frequency cepstral coefficients are base line acoustic features for text independent speaker identification and assumed that Mel frequency cepstral coefficients associated with speakers speech spectra follows a Gaussian Mixture Model.

Sailaja, Sriniva Rao and Reddy (2010a, 2010b, 2010c, 2011)[35][33][32][34] have developed and analyzed text independent speaker identification models with Mel frequency coefficients as feature vectors and follow either Doubly truncated multivariate Gaussian mixture model or Generalized Gaussian mixture model. The Generalized Gaussian mixture model will also include Gaussian mixture model as particular case. In all these papers the authors considered only a first few Mel frequency cepstral coefficients and the remaining coefficients are dropped as insignificant due to high dimensionality problems. They have also assumed that the Mel frequency cepstral coefficients of each speaker speech spectra are independent.

But in reality the adjacent Mel frequency cepstral coefficient are correlated. Ignoring the dependences and dropping some of the Mel cepstral coefficients may lead to a serious falsification of model. So, to have a robust model for the speakers speech spectra, it is need a to reduce the dimensionality and avoid dependence among the features this can be done by further processing the Mel frequency cepstral coefficients of each speakers speech spectra with Independent component analysis.(ICA). The integration of ICA with Mel frequency cepstral coefficients will make the feature vector more robust in characterizing the speech spectra of the speaker.

The ICA, tries to express a set of random variables with some noise as linear combination of components that are statistically independent. ICA is the powerful tool available for high dimensional multivariate analysis. Application of this tool to speech analysis results in computational and conceptual simplicity. Independent component analysis is a technique used for text independent speaker identification system. ICA helps in capturing some of the essential features of speech data in many Speaker identification Systems (Hyvarinen, (2001))[10]. ICA techniques can be used to separate mixed signals. The required assumption in applying ICA is that the observed signals are linearly mixed. As general feature vectors for patter recognition are nonlinearly mixed, distinguishing among characteristics is intractable. However, the normomorphic process of cepstrum parameter extraction, which transforms convolutively mixed filters functions into additive ones, ICA is applied to this problem domain and transformed feature vectors are used for the training and testing of a speaker identification system.

Hence, in this paper a text independent speaker identification method is developed and analyzed integrating independent component analysis (ICA) with Mel frequency cepstral coefficient and using Generalized Gaussian mixture model. The procedure for extracting the feature vectors of the speaker speech spectra using Mel frequency cepstral coefficients and independent component analysis is presented. Assuming that the feature vector of each individual speech spectra follows a Generalized Gaussian mixture model, the estimation of the model parameters is done by using the updated equation of EM algorithm. The speaker identification algorithm under Bayesian frame work is also presented. The efficiency of the developed procedure is studied by conducting experimentation with 50 speakers with 12 utterances of locally recorded data base. A comparative study of different earlier models with the proposed (ICA + MFC coefficients) hybridized method is also discussed.

II. FEATURE VECTORS EXTRACTION

In this section we briefly describe the feature vector extraction of each speaker speech spectra. Reliable and efficient smoothing of the frequency response of a human vocal tract is to be obtained as feature vector. Mel frequency cepstral coefficients have gained importance in speaker identification to describe the signal characteristics. According to psychophysical studies human perception of the frequency content of sounds follows a subjectively defined nonlinear scale called the Mel scale [4] [10], . This is defined as

$$f_{mel} = 2595 \log [1 + (f/700)]$$

Where, f_{mel} is the subjective pitch in Mel's corresponding to f , the actual frequency in Hz. This leads to the definition of MFCC, a base acoustic feature for speech and speaker recognition applications. ICA is the liner and a supervised dimensional reduction algorithm. ICA is not necessarily orthogonal and it extracts independent components even with smaller magnitudes. This section gives brief introduction and analysis of ICA algorithms. Text independent speaker identification system is a highly complex model associated with a huge number of feature vectors. Analysis of such models a challenging task. Under such circumstance dimensional reduction of the data becomes a major requirement for obtaining better identification results. This can be achieved by using ICA algorithms.

Each person's voice has distinguishing properties and features which makes them unique. Air stream pumped by the lungs modifies itself to generate desired sequence of sounds every time a person tries to speak. This implies that there exist some differences in the characteristics of speech depending on the changes in the shape of the vocal tract, vibration of the vocal chords and the nasal cavity. Vocal tract can then be considered as a set of filters that change or alter a set of excitation signals. ICA aims at extracting a set of statistically independent vectors from the matrix of training data, the Mel-frequency Cepstral feature vectors derived from the original signal. It tends to find directions of minimum mutual information. It aims at capturing certain Correlations among the frequencies present in the spectral based representation of a speech signal. This is achieved by ICA in the form of linear combinations of basic filter functions specific to each person. Specific sounds are then generated by combining these functions in a statistically independent nature.

The signal \mathbf{X} is used as a proper Mel-Cepstral based representation of the original signal and the data can be observed as a set of multivariate time series resulting from a hidden linear mixing process \mathbf{A} of independent functions \mathbf{S} (Potamitis, (2000))[25], (Hyvarinen, (2001))[14]. Linear combination of such sources or functions can be summarized as (Cardoso, (1996))[4]

$$\mathbf{Ax} = \mathbf{S} \quad (1)$$

The problem of ICA is to determine both the excitation signal \mathbf{s} and the scalars \mathbf{A} and the only known component is the matrix of the MFCC coefficients of the input speech signal. \mathbf{s} can be computed as follows (Hyvarinen, (1997))[16]

$$\mathbf{S} = \mathbf{A} \mathbf{x}^{-1} \quad (2)$$

\mathbf{A} can be computed by consider \mathbf{x} as a vector of observations, where each observation is expressed as a linear combination of independent components. In order to estimate one of the independent components, a linear combination of x_i is chosen such that (Hyvarinen, (1997))[16], (Michael, (2002))[19]

$$\mathbf{y} = \mathbf{W}^T \mathbf{x} \quad (3)$$

with respected to the conditions stated in equation (1) and equation (2), the linear combination represents in equation (3) is a true combination of independent components if w_i were one of the columns of the inverse of \mathbf{A} . After preprocessing and whitening, the final equation is given as (Michael, (1999)) [24]

$$S \approx \mathbf{y} = \mathbf{W}^T \mathbf{X} = \tilde{\mathbf{W}} \mathbf{P} \mathbf{x} \quad (4)$$

Fast ICA algorithm (Hyvarinen, (1999))[11] is used to estimate w_i which constitutes the rows of \mathbf{W} . Since the components are considered to be statistically independent, the variance between them is high. The following steps are used to estimate w

1. Choose an initial random guess for \mathbf{W}
2. Iterate: $\mathbf{W} \leftarrow E\{xg(W^T X)\} - E\{g'(W^T X)\}W$
3. Normalize: $\mathbf{W} \leftarrow \frac{w}{\|w\|}$
4. If not converged, go back to step 2.

Once w is estimated the final step is to project the signal into the space created by ICA.

New dataset = $\mathbf{W}_{ica} * \text{Mean Adjusted original Data}$,

where, \mathbf{W}_{ica} is the transformation matrix obtained from Fast ICA algorithm.

The multivariate dimensionality reduction techniques (ICA) can be applied to Mel spectral energies (Ding, (2001))[9], or the Mel frequency cepstral coefficient after the feature vectors obtained through MFCC's makes the dimension reduction possible and more efficient (Wanfeng, 2003)[36]. This is because the Mel frequency feature vector characteristics agree with the assumptions made in ICA algorithms (Somervuo, (2003))[30].

The extraction of Feature vectors for speaker identification is done in two steps. :

1. Compute MFCC's, and 2. Apply ICA to transform them to get new feature vectors.

The computation steps for extracting the new feature vectors are as follows.

Step 1.

- (a) Take the Fourier transform of a signal
- (b) Map the powers of the spectrum obtained above on to the Mel scale, using triangular overlapping windows
- (c) Take the logs of the powers at each of the Mel frequencies

- (d) Take the discrete cosine transform of the list of the Mel log powers, as it were a signal
- (e) The MFCCs are the amplitudes of the resulting spectrum.

Step 2.

- (f) Apply ICA transform to Mel frequency cepstral coefficients to get the new feature vectors of each speaker speech spectra

III. SPEAKER IDENTIFICATION MODEL WITH GENERALIZED GAUSSIAN DISTRIBUTION:

In this section we describe the speaker identification process. Fig.1 represents the block diagram of the proposed text independent speaker identification system with Generalized Gaussian distribution using integrating ICA in the system after feature extraction.

Here it is assumed that the feature vector (after processing the MFCC with ICA) follows a multivariate Generalized Gaussian mixture model. The motivation for considering the Generalized Gaussian mixture models is that the individual component densities of a multi model density like the mixture model may model some underlying set of acoustic processes.

It is reasonable to assume the acoustic space corresponding to a speaker voice can be characterized by acoustic classes representing some broad phonetic events such as vowels, nasals or fricatives. These acoustic classes reflect some general speaker dependent vocal tract configurations that are useful for characterizing speaker identity. The spectral shape of its acoustic class can in turn be represented by the mean of its component density and the variation of the average spectral shape can be represented by the co-variance matrix. Therefore the entire speech spectra of the each individual speaker can be characterized as a M component Finite Multivariate Generalized Gaussian mixture distribution.

The probability density function of the each individual speaker speech spectra is

$$p(\vec{x}_t | \lambda) = \sum_{i=1}^M \alpha_i b_i(\vec{x}_t | \lambda) \quad (2.1)$$

where, $\vec{x}_t = (x_{tj})$ $j=1,2,\dots,D$; $i=1,2,3,\dots,M$; $t=1,2,3,\dots,T$ is a D dimensional random vector representing the MFCC vector

λ is the parametric set such $\lambda = (\mu, \rho, \Sigma)$

α_i is the component weight such that $\sum_{i=1}^M \alpha_i = 1$

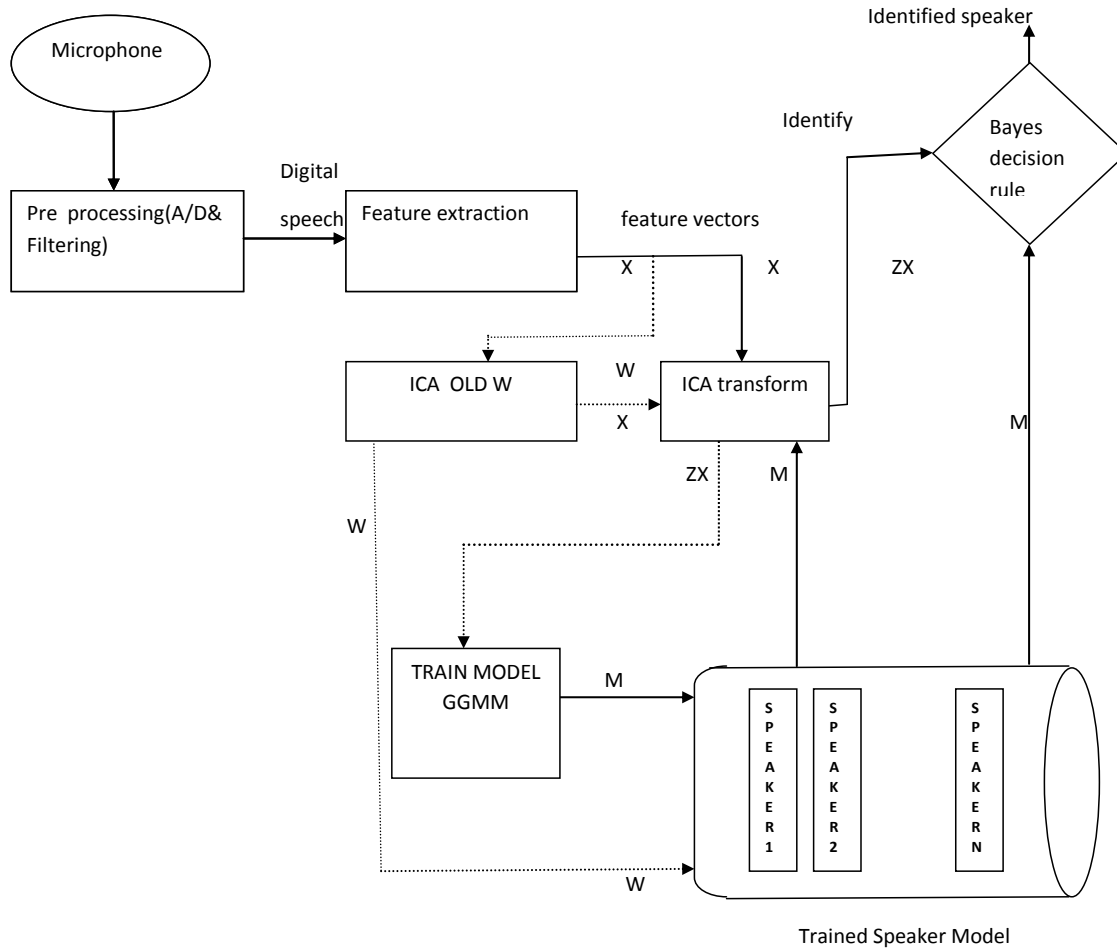


Fig 1. Schematic diagram representing the Speaker Identification process

$b_i(\vec{x}_t|\lambda)$ is the probability density of i^{th} acoustic class represented by new vectors of the speech data and the D -dimensional Generalized Gaussian (GG) distribution (M. Bicego et al (2008))(23) and is of the form

$$b_i(\vec{x}_t|\lambda) = \frac{[\det(\Sigma)]^{-1/2}}{[z(\rho)A(\rho,\sigma)]^D} \exp\left(-\left\|\frac{\frac{1}{z}(\vec{x}_t - \vec{\mu}_i)}{A(\rho,\sigma)}\right\|_{\rho}\right) \quad (2.2)$$

Where, $z(\rho) = \frac{2}{\rho} \Gamma\left(\frac{1}{\rho}\right)$ and

$$A(\rho, \sigma) = \sqrt{\frac{(1/\rho)}{(3/\rho)}} \quad (2.3)$$

And $\|x\|_{\rho} = \sum_{i=1}^D |x_i|^{\rho}$ stands for the l_{ρ} norm of vector x , Σ is a symmetric positive definite matrix. The parameter $\vec{\mu}_i$ is the mean vector, the function $A(\rho)$ is a scaling factor which allows the $\text{var}(x) = \sigma^2$ and ρ is the shape parameter when $\rho=1$, the Generalized Gaussian corresponds to a laplacian or double exponential Distribution. When $\rho=2$, the Generalized Gaussian corresponds to a Gaussian distribution. In limiting case $\rho \rightarrow +\infty$ Equation (2.2) Converges to a

uniform distribution in $(\mu - \sqrt{3}, \mu + \sqrt{3})$ and when $\rho \rightarrow 0+$, the distribution becomes a degenerate one when $x = \mu$.

The model can have one covariance matrix per a Generalized Gaussian density of the acoustic class of each speaker. Based on the previous studies, the diagonal covariance matrix is used for speaker model. As a result of diagonal covariance matrix for the feature vector, the features are independent and the probability density function of the feature vector is

$$b_i(\vec{x}_t|\lambda) = \prod_{j=1}^D \frac{\exp\left(-\left|\frac{(x_{tj} - \mu_{ij})}{A(\rho_{ij}, \sigma_{ij})}\right|^{\rho_{ij}}\right)}{\frac{2}{\rho_{ij}} \Gamma\left(1 + \frac{1}{\rho_{ij}}\right) A(\rho_{ij}, \sigma_{ij})} = \prod_{j=1}^D f_{ij}(x_{tj})$$

To find the estimate of the model parameters μ_{ij} and ρ_{ij} for $i=1,2,3 \dots, M, j=1,2, \dots, D$, we maximize the expected value likelihood (or) log likelihood function. Here the shape parameters ' ρ_{ij} ' is estimated by the procedure given by Armando.J el at (2003) [1] for each acoustic class of each speech spectra.

The updated equations of the parameters for EM algorithm are as given by Sailaja et al (2010) [34] are

The updated equation for estimating α_i is

$$\alpha_i^{(l+1)} = \frac{1}{T} \sum_{t=1}^T \left[\frac{\alpha_i^{(l)} b_i(\vec{x}_t, \lambda^{(l)})}{\sum_{i=1}^M \alpha_i^{(l)} b_i(\vec{x}_t, \lambda^{(l)})} \right]$$

Where $\lambda^{(l)} = (\mu_{ij}^{(l)}, \sigma_{ij}^{(l)})$ are the estimates obtained

The updated equation for estimating μ_{ij} is

$$\mu_{ij}^{(l+1)} = \frac{\sum_{t=1}^T t_1(\vec{x}_t, \lambda^{(l)}) A(N, \rho_{ij}) (x_{tj} - \mu_{ij})}{\sum_{t=1}^T t_1(\vec{x}_t, \lambda^{(l)}) A(N, \rho_{ij})}$$

where, $A(N, \rho_{ij})$ is some function which must be equal to unity for $\rho_i = 2$ and must be equal to $\frac{1}{\rho_{ij}-1}$ for $\rho_i \neq 1$, in the case of $N=2$, we have also observed that $A(N, \rho_{ij})$ must be an increasing function of ρ_{ij} .

The updated equation for estimating σ_{ij} is

$$\sigma_{ij}^{(l+1)} = \left[\frac{\sum_{t=1}^N t_1(\vec{x}_t, \lambda^{(l)}) \left(\frac{\Gamma(\frac{3}{\rho_{ij}})}{\rho_{ij} \Gamma(\frac{1}{\rho_{ij}})} \right) |x_{tj} - \mu_{ij}^{(l)}|^{\frac{1}{\rho_{ij}}}}{\sum_{t=1}^T t_1(\vec{x}_t, \lambda^{(l)})} \right]^{\frac{1}{\rho_{ij}}}$$

IV. SPEAKER IDENTIFICATION USING BAYES' DECISION RULE

The extracted feature vectors from each test speaker were applied to the function "ICA Transform" (Fig. 1) and were estimated into the space of ICA created by the associated speaker with unique speaker ID. This uses the stored ICA W from the trained speaker model. The new feature vectors of the test utterances and the trained models were fed to a Byes classifier for identification applications which employ large group of data sets and the corresponding test speaker was identified (Domingo's, (1997))[8][9].

$p(i/x_t, \lambda)$ is called a posteriori probability for an acoustic class i and is defined by the following equation

$$p(i/x_t, \lambda) = \frac{p_i b_i(x_t^{\wedge})}{\sum_{k=1}^M p_k b_k(x_t^{\wedge})}$$

For a given observation sequence the main goal is to find the speaker model that has the maximum posteriori probability represented as (Reynolds,(1995))[7]

$$\begin{aligned} \hat{s} &= \max_{1 < k < S} p_l(\lambda_k | X) \\ &= \arg \max_{1 < k < S} [p(\lambda_k | X) p_r(\lambda_k)] \end{aligned}$$

The speaker identification system finally computes S using the logarithms and the independence between the observations.

V. EXPERIMENTAL RESULTS

The performance of the developed model is evaluated by using a database of 50 speakers. For each speaker there are 12 conversations of approximately 4sec.each recorded in 12 separate sessions. Out of which three, four & five sessions are used for training data and the remaining sessions used for testing data. The speaker's speech data was recorded locally by using high quality Microphone. The test speech was first processed by front end analysis explained in section 2 to produce a new transformed feature vectors which are obtained for test sequence length 4 seconds, ith the procedure given by Smitha Gangisetty (2005)[29].

Using the classified data for each speaker the updated equations of the parameters are calculated by the procedure explained in section 3, the refined estimates of the parameters are obtained. The global model for each speaker density is estimated by using the derived parameters. With the test data set, the efficiency of the developed model is studied by identifying the speaker with the Speaker identification algorithm given in section (4). The average percentage of identification was computed and the results are tabulated. Two sets of experiments were carried out for evaluating the performance of the develop model.

EXPERIMENT -1. This experiment involves clean train and test signals.

TABLE 1. AVERAGE PERCENTAGE OF CORRECT IDENTIFICATION VERSUS FOR VARIOUS SPEAKER IDENTIFICATION MODELS

Model	Percentage of accuracy
GMM-nv	94.6±1.8
GMM-gv	89.7±2.4
TGMM	80.2±3.1
GC	67.3±3.76
FDTMGMM (K-means)	96.4±1.7
FDTMGMM (Hierarchical clustering)	97.1±1.
Embedded ICA with GMM	94.12%
Embedded ICA with GGMM	97.8±12.

A comparative study of the Performance of the developed model is carried with reference to existing speaker modeling techniques. Specially the other techniques are the unimodel Gaussian classifier given by H.Gish (1985)[12], Tied Gaussian Mixture model given by J. Oglesby and J. Mason,(1991)[20] and the Gaussian Mixture Model using nodal variance(GMMnv) and Gaussian Mixture Model using global variance (GMMgv) by Douglas A Reynolds (1994)[5], and Finite Doubly Truncated Gaussian Mixture Model .

From, Table 1, it is observed that the average percentage of correct identification for the developed model is

97.8±12 The percentage correctness for the Gaussian Mixture Model with integrated ICA is 94.12%. This clearly shows that the speaker identification model with multivariate Generalized Gaussian Mixture Model is having higher average percentage of correct identification than the other models.

EXPERIMENT -2. Additive white Gaussian noise (AWGN) affected train signals at a particular SNR of 30dB and AWGN affected test signals with varying SNR's. From the table we found that as the signals to noise ratio increases the identification rate also increases. The performance of speaker identification is improved using ICA with Generalized Gaussian Mixer Model even in noisy conditions.

TABLE 2. PERFORMANCE OF PCA AND ICA WITH VARIATION IN SNR OF THE TEST SIGNALS

MODEL TRAIN : 30dB SNR	TEST VALUES OF SNR dB		
	0	10	20
ICA with GMM	51.00%	70.00%	85.50%
PCA with GMM	40.33%	62.41%	76.70%
ICA with GGMM	55.01%	72.34%	88.45%

VI. CONCLUSION

In this paper a novel Text Independent Speaker Identification model is developed with the assumption that the feature vector associated with the speech spectra of each individual speaker follows a Generalized Gaussian Mixture Model. The Generalized Gaussian Mixture Model also includes the leptokurtic or platykurtic nature of the feature vector associated with each vocal class of individual speaker's speech spectrum. The new feature vectors are derived for the each speaker's speech data through the mel frequency cepstral coefficients + ICA.

This procedure avoids the losing of information on speech spectra, by dropping some of the MFC Coefficients and avoids the dependencies among them. The model parameters are obtained by deriving the updated equations from the EM Algorithm associated with Finite Multivariate Generalized Gaussian Mixture Model. An experimentation with 50 speakers speech data revealed that this Text Independent Speaker Identification with integrating ICA using Finite Multivariate Generalized Gaussian Mixture Model outperforms the earlier existing Text Independent speaker identification models. This method works well even with large population with small time speech data sets.

REFERENCES

[1] Armando. J et al (2003), "A practical procedure to estimate the shape Parameters in the Generalized Gaussian distribution."
[2] Ben Gold and Nelson Morgan (2002), "speech and audio processing" part 4 chapter 14 pp.189-203 John willis and sons.
[3] Cardoso(1996) Automated speech/speaker recognition over Digital wireless 29, pp. 641-662.

[4] Cardoso, (1996), "Equivariant adaptive source separation" IEEE Transaction on signal processing, Volume.44,no.12,pp.3017-3030.
[5] D.A. Reynolds (1994), "Experimental evaluation of features for robust speaker identification", IEEE Trans. Speech Audio Process, pp. 639-643.
[6] D.O. Shoaughnessy (1987), "Speech Communication Human and Machine, wisely publication, NEWYORK.
[7] Douglas A. Reynolds, and Richard C. Rose (1995), "Robust Text Independent Speaker Identification using Gaussian Mixture Speaker Model," IEEE trans. Speech and Audio Processing, vol.3, pp. 72-83.
[8] Domingos, (1997), "On the optimality of the simple Bayesian classifier under zero-one loss," Machine Learning, vol. 29, pp. 103-130, 1997
[9] Domingo(1997), "Speaker Recognition", A Tutorial", Proceedings of the IEEE, Vol. 85, no. 9,
[10] Ding(2001), "Personal Recognition Using ICA," 8th International Conference on Neural Information Processing,
[11] Erell and MBH Juang, LR Rabiner, and JG Wilpon, "On the use of bandpass filtering in speech recognition," IEEE Trans. Acoust., Speech, Signal Processing, vol.573-576.
[12] Furui, (1992), "toward robust speech recognition and adverse conditions" proceeding of the ESCA workshop in speech processing under adverse conditions, PP.31-41.
[13] H. Gish et al (1985), "Investigation Of Text-dependent Speaker Identification Over Telephone Channels," in Proc. IEEE ICASSP, pp. 379-382.
[14] Hyvarinen, (2001), "Independent component analysis, John Wiley and sons.
[15] Hyvarinen,(1999), "Fast and Robust Fixed Point Algorithm for ICA" proceedings of the IEEE transactions on neural networks, Volume.10,no.3,pp.626-634.
[16] Hyvarinen.(1997) "A family of fixed point algorithms for independent component analysis" proceedings of the IEEE international conference on acoustics speech and signal processing (ICASSP), VOLUME5, pp.3917-3920.
[17] J.F. Cardoso (1997), "Informax and maximum likelihood for Blind Source Separation", IEEE Signal Processing Letters, Vol.4.
[18] Jyothi et al (2011), "Text independent speaker identification with finite multivariate generalized Gaussian mixture model with distant microphone speech" proceeding of the international journal of computer applications(IJCA)14(4):5-9,
[19] Heidelberg (2003), "ICANN/ICONIP'03 of the 2003 joint international conference on Artificial neural networks/neural information processing".
[20] J.Oglesby and J. Mason (1991), "Radial basis function networks for Speaker Recognition," in Proceedings of IEEE ICASSP, pp. 393-396.
[21] JP Campbell (1997), "Speaker Recognition: A Tutorial", Proceedings of the IEEE, Vol. 85, no.9.
[22] Leena Mary and Yegnanarayana(2008), "Extraction and representation of prosodic feature for language and speaker recognition" SPEECH COMMUNICATION 50(10):782-796. Michael Charles (1999), "Orthogonal GMM in Speaker Recognition," Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing", pp. 845-848.,
[23] Md M. Bicego, D Gonzalez, E Grosso and Alba Castro(2008) "Generalized Gaussian distribution for sequential Data Classification" IEEE Trans. 978-1-4244-2175-6.
[24] Ning Wang, P. C. Ching(2011), "Nengheng Zheng, and Tan Lee, "Robust Speaker Recognition Using Denoised Vocal Source and Vocal Tract Features speaker verification," IEEE Transaction on Audio Speech and Language processing, Vol. 1, No. 2, pp. 25-35.
[25] Potamitis, N. Fakotakis and G. Kokkinakis (2000), "Independent component analysis applied to feature extraction for robust automatic speech recognition", IEE Electronic Letters, Vol.36, No.23, pp.1977-1978. Pierre C (1994). "Independent component analysis", A new concept? Signal Processing, 36:pp. 287-314.
[26] Sandipan Chakraborty, Anindya Roy and Goutam Saha(2006) "Improved Closed Set text-Independent Speaker identification by combining MFCC with Evidence from Flipped Filter banks " IJSp vol 4, no 2
[27] S. Furui (1986), "Cepstral analysis technique for automatic speaker verification", IEEE Trans. Acoust. Speech Signal Process., Vol. 29, No. 2, pp. 254-272.

- [28] S. Furui(1982), "Speaker recognition by statistical features of Speech Communication S. Furui / Individual features in speech waves /97 cepstral parameters", Trans. IECE, Vol. J65-A, pp. 183-190.
- [29] Smitha Gangisetty (2005), "Text-Independent Speaker Recognition", MS Thesis, College of Engineering and Mineral Resources at Morgantown, West Virginia University.
- [30] Somervuo, chen B and Zhu Q, ,(2003) "Feature Transformations and Combinations for Improving ASR Performance" proceedings of the 8th European conference on speech communication and technology (Euro speech),PP.477-480.
- [31] Somervuo,(2003) " Experiments with Linear and Non Linear Feature Transformations in HHM based phone Recognition" proceedings of the IEEE International Conference on Acoustics, Speech and signal processing (ICASSP), volume.1,PP.52.55.
- [32] V Sailaja, K Srinivasa Rao & K V V S Reddy(2010), "Text Independent Speaker Identification Using Finite Doubly Truncated Gaussian Mixture Model " International Journal of Information Technology and Knowledge Management July-December 2010, Volume 2, No. 2, pp. 475-480
- [33] V Sailaja, K Srinivasa Rao & K V V S Reddy(2010), "Text Independent Speaker Identification Model Using Finite Doubly Truncated Gaussian Distribution And Hierarchical Clustering" International Journal of Computer Science & Communication Vol. 1, No. 2, , pp. 333-338.
- [34] V Sailaja, K Srinivasa Rao & K V V S Reddy(2010), "Text Independent Speaker Identification Model with Finite Multivariate Generalized Gaussian Mixture Model and Hierarchical Clustering Algorithm " International Journal of Computer applications Vol. 11, No. 11, , pp. 25-31.
- [35] V.Sailaja (2010), "Some Studies on Text Independent Speaker Identification Models with Generalizations of Finite Gaussian Mixture Models", unpublished Thesis notes Department of Electronics and Communication Engineering, Andhra University, Visakhapatnam.
- [36] Wanfeng,(2003) "experimental evaluations of a new speaker identification frame work using PCA" IEEE international conference on systems ,man and cybernetics, volume.4147-4152,PP.5-8.
- [37] Z. Wanfeng et al (2003), "Experimental Evaluation of a New Speaker Identification Framework using PCA", IEEE Inter- national Conference on Systems, Man and Cybernetics, Vol. 5, pp.4147-4152.
- [38] Zhu(1994), "Text-independent speaker recognition using VQ, mixture Gaussian VQ Andergodic HMMs", in ASRIV,pp.55-58.

AUTHORS PROFILE

*Mr N.Murali .Ramalingeswararao Completed M.tech in Digital Electronics and Communication Engineering in GIET in Jawaharlal Nehru Technological University, Kakinada, INDIA in 2011.

* Vemuri Sailaja is professor of electronics and communication engineering at GIET. She received his Ph.D instatistical signal processing from Andhra University. She published 10 research papers in referred International and National Journals and She guided 10 M.Tech students. She is a fellow of IETE. She also presents several papers in National and International conferences.

*Dr.K.S.Rao is professor of statistics at Andhra University. He Introduced four new and novel probability distributions namely bimodal distribution, generalized Laplace distribution, four parameters generalized Gaussian distribution and published. He received his Ph.D degrees and he published 88 research papers in refereed international and national journals. He completed 9 research projects sponsored by ministry of industries, Govt of India of AP, UGC and vizag airport. He was the chief editor of Journal of Indian Society for probability and statistics.vol12.

Energy Efficient Zone Division Multihop Hierarchical Clustering Algorithm for Load Balancing in Wireless Sensor Network

Ashim Kumar Ghosh¹, Anupam Kumar Bairagi², Dr. M. Abul Kashem³, Md. Rezwan-ul-Islam¹, A J M Asraf Uddin¹

¹Department of CSE, DUET, Gazipur-1700, Bangladesh

²Lecturer, Department of CSE, KU, Khulna-9208, Bangladesh

³Associate Professor, Department of CSE, DUET, Gazipur-1700, Bangladesh

Abstract— Wireless sensor nodes are use most embedded computing application. Multihop cluster hierarchy has been presented for large wireless sensor networks (WSNs) that can provide scalable routing, data aggregation, and querying. The energy consumption rate for sensors in a WSN varies greatly based on the protocols the sensors use for communications. In this paper we present a cluster based routing algorithm. One of our main goals is to design the energy efficient routing protocol. Here we try to solve the usual problems of WSNs. We know the efficiency of WSNs depend upon the distance between node to base station and the amount of data to be transferred and the performance of clustering is greatly influenced by the selection of cluster-heads, which are in charge of creating clusters and controlling member nodes. This algorithm makes the best use of node with low number of cluster head know as super node. Here we divided the full region in four equal zones and the center area of the region is used to select for super node. Each zone is considered separately and the zone may be or not divided further that's depending upon the density of nodes in that zone and capability of the super node. This algorithm forms multilayer communication. The no of layer depends on the network current load and statistics. Our algorithm is easily extended to generate a hierarchy of cluster heads to obtain better network management and energy efficiency.

Keywords- routing protocol; WSN; multihop; load balancing; cluster based routing; zone division.

I. INTRODUCTION

WSNs consists of more than hundreds of small spatially distributed autonomous devices using sensor called sensor nodes to monitor the physical and environmental situations such as sound vibration, temperature, pressure, motion and intensity of light at various place. Energy is most concentrate term in WSNs because it determines the aliveness of wireless sensor node. One of the most design objectives of WSNs is to minimize node energy consumption and maximize the network life time [1][2]. So preserving the consumed energy of each node is an important objective that must be considered when developing a routing algorithm for wireless sensor networks. In WSNs each node tries to perform computation on data locally, so data to be forwarded condenses because computations is less expensive than data transmission in WSNs. e.g. to calculate the mean value of data sample at node

is much efficient than to transmit sample data and calculate the mean value at base station. Due to the short range of the radio communication and the fact that consumption of energy is proportional to the square of the distance making communication multi hop instead of direct communication will save energy.

The rest of this paper is prepared as follows segment II briefly describes the applications of the WSN in different areas, Segment III describes the problem formulation, Segment IV gives general key issues of the various routing algorithms, Segment V includes a detailed study of the related research. The proposed algorithm is discussed in segment VI, segment VII discusses the simulation and its results and lastly concludes the paper.

II. APPLICATION

Here we describe a few areas where WSNs can be used effectively. According to [3] WNs are able to supervise wide range of applications which include Intensity of light, Temperature, Humidity, Pressure, existence of objects, size and motion of objects. Usually applications include inspection and battle space monitoring [4] by military, agricultural and environmental. Engineering applications include maintenance in a large industrial plant, regulation of modern buildings in terms of temperature, humidity etc. Other applications include greenhouse monitoring, land slide detection, forest fire detection, flood detection etc. [5].

III. PROBLEM FORMULATION

Broadcasting is the process in which a source node sends a message to all other nodes in the network that's deal with energy [18]. The energy factors lead to the cost increase of WSN continuation, specifically for the networks deployed in out-of-the-way areas.

Thus, the problem of WSN long-term operation is of vital significance. The lifetime problem is a complex problem of WSN and cannot be resolved one-sidedly. The problems have to be introduced at three main categories [8][15][16]:

Design - application, modeling, simulation;

Hardware - hardware machinery, technology, maximum power point tracking, energy scavenging technology;

Software - energy saving techniques, communication protocol, middleware.

At the design level the full network operation depends on the application. This application may decide the hardware architecture and power management approach of the sensor node and network. As well, earlier to hardware implementation of the node it should be modeled and the full network has to be simulated.

The lifetime problem at the hardware level includes the correct HW machinery choice along with energy storage. Energy scavenging technology may significantly increase the WSN lifetime, but for the maximum efficiency it is desirable to apply it with the maximum power point tracking (MPPT) technique [6].

Software plays a vital responsibility in WSN lifetime as well. e.g. the adaptive duty-cycling algorithm [7] allows the utilization of up to 58% more environmental resources in comparison to the systems without this technique.

IV. DESIGN ISSUE OF ROUTING PROTOCOL

WSNs pose various challenges to design. This segment summaries some of the major challenges faced while clustering the wireless sensor network.

A. Network scalability

The number of sensor nodes deployed in the sensing area may be in the order of hundreds, thousands or more. When a WSN is deployed, sometime new nodes need to be added to the network, so routing algorithm must be scalable enough to respond to actions.

B. Mobility

Wireless node have the propriety of mobility, Mobility of nodes is an significant issue in mobile ad-hoc networks (MANET). Nodes in MANET move from one network to another separately or in group. In single node mobility scheme every node performs registration individually in new MANET whereas in group mobility method only one node in a group.

C. Network deployment

Node deployment in WSNs is either fixed or random depending on the purposes where it will be used. In fixed deployment the network is deployed on preset locations whereas in random deployment the resulting distribution can be uniform or not uniform. In such a case careful managing the network is necessary in order to guarantee full area coverage and also to guarantee that the energy consumption is also uniform across the network.

D. Multihop or Singla hop communication

The communication model of wireless sensor network may be formed as single hop or multi hop. Energy consumption in wireless systems is directly proportional to the square of the distance; single hop communication is costly with respect to energy consumption. Most of the routing algorithms use multi hop communication model since it is more energy efficient

with respect to energy consumption in contrast, with multi hop communication the nodes which are nearer to the cluster head.

E. Cluster dynamics

Cluster dynamics describes how the unlike parameters of the cluster are determined e.g., the number of clusters in a particular network. In some cases the number might be pre assigned and in some cases it is dynamic. In case of dynamic, there is an option of forming unbalanced clusters. While limiting it by some pre-assigned, minimum distance can be effective in some cases. Also cluster head selection can either be centralized or decentralized which both have advantages and disadvantages. The number of clusters might be fixed or dynamic. Fixed number of clusters cause less overhead in that the network will not have to repeatedly go through the set up phase in which clusters are formed.

V. RELATED WORKS

Routing in WSNs is a tricky task because of data source from multiple paths to single source, data redundancy and also because of energy and computation factors of the network [9]. The usual routing algorithms are not efficient when applied to WSNs. The performance of the existing routing algorithms for WSNs varies from application to application because of various demands of various applications.

Generally the routing protocols are classified into two classes one is based on the network structure and the second is based on protocol operation. The network structures are further classified as flat network routing, hierarchal network routing and location based routing. The protocol operation can be classified as negotiation based, query based, multipath based, coherent based and QoS based routing. Rest of section shortly describes the routing protocols based on network structure and more specifically the hierarchal routing algorithms.

Cluster based routing in WSNs comes under the class of hierarchal routing. Hierarchal routing comprise the formation of clusters where nodes are assigned the task of sensing which have low energy and transmission task to nodes which have higher energy. The cluster heads may be extraordinary nodes with higher energy or ordinary node depending on the algorithm and application. The cluster head also performs computational functions for instance data compression and data aggregation in order to decrease the number of transmission to the base station in this manner saving energy of that node. Load balancing and redistribution (LBR) technique for adhoc networks in general and wireless sensor networks in particular cases have been described in [19]. A Genetic Algorithm (GA) [20] is used to generate energy-efficient hierarchical clusters. Static clustering techniques [21] and MTE routing is also increase the efficiency of the network life time. The multi-hop technique with a backoff [22]-based clustering algorithm to organize sensors. By using an adaptive backoff strategy, the algorithm not only realizes load balance among sensor node, but also achieves fairly uniform cluster head distribution across the network.

Low Energy Adaptive Clustering Hierarchy (LEACH) is the first hierarchical cluster-based routing protocol for wireless sensor network which divided the nodes into clusters,

in each cluster a dedicated node with extra privileges called Cluster Head (CH) is responsible for creating and manipulating a TDMA (Time division multiple access) schedule and sending aggregated data from nodes to the BS where these data is needed using CDMA (Code division multiple access). Rest nodes are cluster members [10].

LEACH cluster head is chosen using a threshold $T(n)$

$$T(n) = \begin{cases} \frac{p}{1 - p(r \bmod (1/p))} & n \in G \\ 0 & \text{others} \end{cases}$$

Where, p is the percentage of cluster heads over all nodes in the network, r is the number of rounds of selection, and G is the set of nodes that are not selected in round $1/p$.

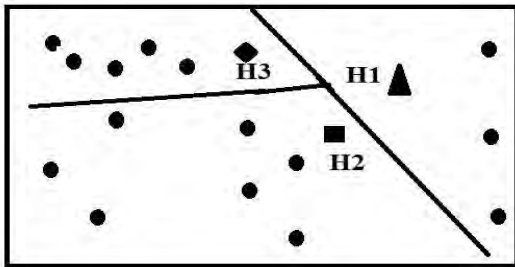


Figure 1. Randomly generated Cluster head using threshold function

WSNs are autonomous networks. Sensor nodes are independent with each other. The coordination between nodes is done through wireless communication, which costs much. This is one of the major reasons that the LEACH protocol selects cluster heads randomly. In Figure, H1, H2 and H3 are three cluster heads; symbolizes as a nodes \blacktriangle , \blacksquare and \blacklozenge respectively. H1, H2 and H3 are very closely located. According to data communication model, the energy that a cluster head consumes is the sum of that consumed in receiving data and that in sending data, as described in equation:

$$E_{ch} = IE_{elec} N_{mem} + IE_{DA} (N_{mem} + 1) + IE_{elec} + I_{emp} d_{toBS}$$

Lin SHEN and Xiangquan SHI [11] describes in there paper, when multiple cluster heads are randomly selected within a small area, a big extra energy loss occurs. The amount of lost energy is approximately proportional to the number of cluster heads in the area. When perform long time, it will experience unbalanced remaining energy among nodes. Blind nodes will appear and the network life and routing efficiency will be decreased.

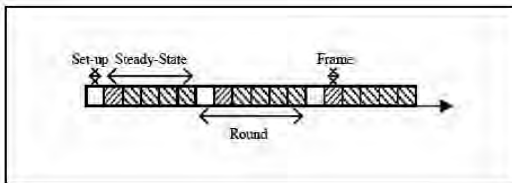


Figure 2. Phases of LEACH Protocol [12]

The operation of LEACH is broken up into *rounds*, where each round begins with a set-up phase, when the clusters are organized, followed by a steady-state phase, when data transfers to the base station occur. In order to minimize overhead, the steady-state phase is long compared to the set-up phase. The phase of LEACH protocol is divided into two sub groups:

Set-up Phase: (1) Advertisement Phase (2) Cluster Set-up Phase

Steady Phase: (1) Schedule Creation (2) Data Transmission

Since LEACH has many shortcomings, many researchers have been done to improve this protocol. Loscri [17] proposed two-level Leach. In this protocol, CH collects data from other cluster members as original LEACH, but rather than transfer data to the BS directly, it uses one of the CHs that lies between the CH and the BS as a relay station. Some improve version of LEACH are as follows: E-LEACH, TL-LEACH, M-LEACH AND V- LEACH etc [14].

VI. ZONE DIVISION MULTI HOP HIERARCHICAL CLUSTERING FOR LOAD BALANCING

Our proposed algorithm consists of two different phases, the setup phase and steady phase. During the setup phase Super nodes known as Cluster Heads and Vice super node as known as temporary cluster heads are chosen followed by the steady phase. The steady phase is the data transmission phase and is longer than the setup phase. In the setup phase, initially the algorithm divide the full region into four zone and also find the centre area of that region, make a set of temporary super node and also select a node super node . The zone may be or not divided again that's depending on efficiency of super node.

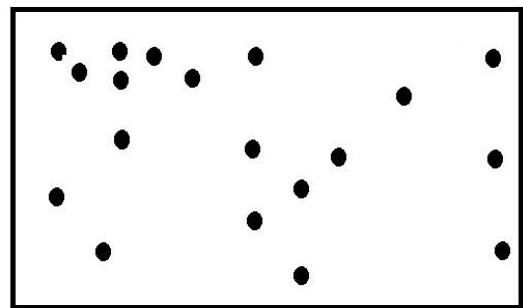


Figure 3. Random deployment of node in network

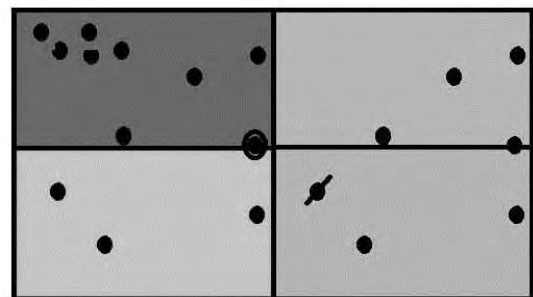


Figure 4. Top layer

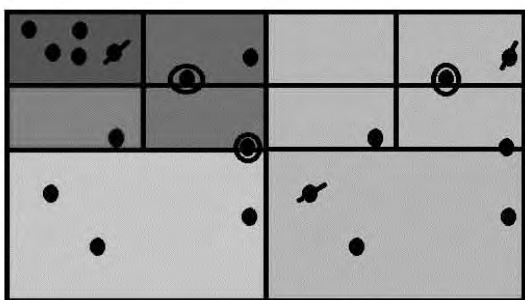


Figure 5. Middle layer

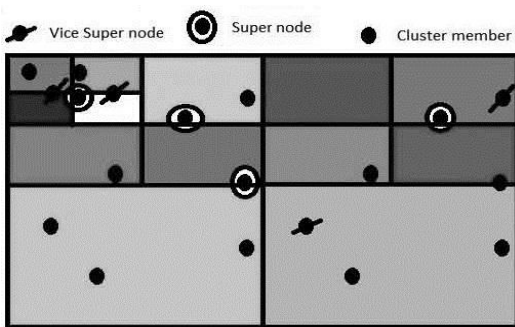


Figure 6. Bottom layer

Figure 3 shows the random deployment of nodes in a network; Figure 4, Figure 5 and Figure 6 shows how cluster are formed with super mode and vice super node by zone division. Zone division helps us to make cluster balanced.

When the super node will be dead the vice super node act as a super node. After finishing the setup phase the steady state phase will start and nodes transmit data. When all the nodes within the cluster finish sending data the super nodes performs some computation on it and sends it to base station using multihop communication.

PROPOSED ALGORITHM:

Setup phase:

1. Measure the region and find the centre of region.
2. Find nodes as close as centre is called set of super node, store the set and specify a super node from the set on the basis of energy level of the node.
3. ID of super node is stored in to the table of previous super node and vice versa.
4. This super node broadcast a message to the all node of that zone and receives **reply message [a]** from the node.
5. If the location or distance and no of node is more than the **efficiency [b]** of super node then divide the region into four (4) equal Zones and go to step 6.

Else

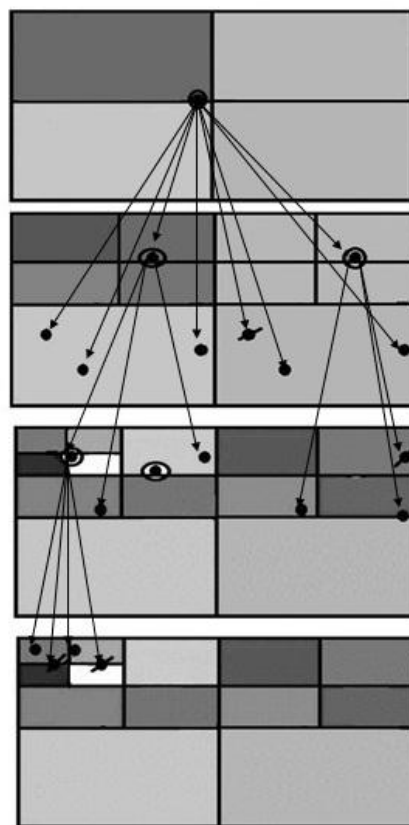


Figure 7. Hiarchical Structure

go to step 7.

6. Find the centre of zone and repeat step 2.
7. The node id is stored in the super node and vice verse,(The super node sends a message about the information of all neighbor nodes of that zone to the node.

Or

Node sends a hello signal to the neighbor nodes.)
Update the neighbors table.

Steady state Phase:

8. **Node to super node communication:** Nodes sensing and transmitting data to the immediate super node in their allotted time slot. The super node collects data and processes the data. After that the super node transmission is start. All super nodes do the same task.

When the data reach to base station **the steady state is repeated.**

a: Reply message contains the energy level.

b: Efficiency of super node is measured by the super node energy level ,signal receiving time and delay of access .

VII. SIMULATION AND RESULT

We calculate the performance of the proposed algorithm in this segment using OMNeT++ simulator. It has been developed by András Varga [13]. It is an object-oriented modular discrete event network simulator. It model consists of hierarchically nested modules. Modules can have their own parameters. Parameters can be used to modify module activities and to parameterize the model's topology.

In our case the sample deployment of the network is shown in Figure 2, Figure 3 shows the top layer of the network with super nodes, Figure 4 shows the middle layer and Figure 5 shows the bottom layer of the network. After formation of layers, clusters are formed and steady state phase starts.

TABLE I. DADED NODE COMPARISON

Time	No of Dead Nodes		
	Proposed algorithm	VLEACH	LEACH
5	3	7	9
10	6	8	13
15	6	10	18
20	7	11	22
25	8	11	25
30	9	13	27
35	11	16	29
40	12	17	30

TABLE II. REMAINING ENERGY CONSUMPTION COMPARISON

Time	Remaining Energy		
	Proposed algorithm	VLEACH	LEACH
5	295	292	255
10	293	240	215
15	280	230	205
20	272	226	180
25	270	222	173
30	263	200	150
35	254	175	143
40	249	171	135
45	241	168	131

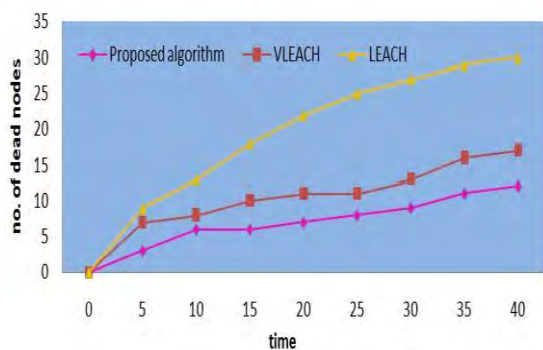


Figure 8. Dead node comparison in network (w.r.t TABLE I)

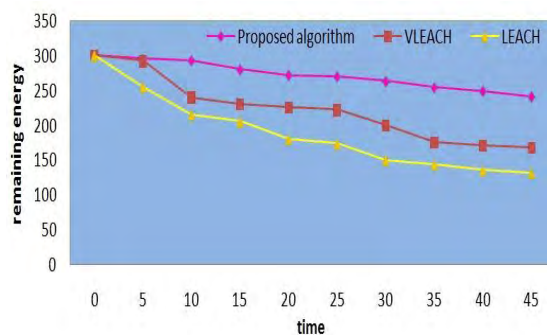


Figure 9. Remaining energy comparison in network (w.r.t. TABLE II)

VIII. CONCLUSION AND FUTURE WORK

It is clear from the simulation outcome that by utilizing the density property of the WSNs it is possible to enhance the network life time and also efficiently balance the energy consumption load across the network. The energy consumption of the network becomes uniform. That mean the proposed algorithm is efficient than the other version of LEACH protocol. The future work includes network with some mobility in the network.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for the helpful comments and suggestions.

REFERENCES

- [1] L. Sun, J. Li, Y. Chen and H. Zhu, "Wireless Sensor Networks," Tsinghua University Publishers, May 2005.
- [2] Ren, H. Huang and C. Lin, "Wireless sensor networks," *Journal of Software*, 2003, 14(7):1282-1291..
- [3] Akyildiz I. F., Su W, Sankarasubramanian Y, Cayirci E, 2002"Wireless sensor network: a survey" *Computer Networks*, 393-422.
- [4] Haenggi M., 2005"Opportunities and Challenges in Wireless Sensor Network," in *Handbook of Sensor Networks Compact wireless and Wired Sensing Systems*: CRC press, 21-34.
- [5] Bonnet P., Seshadri P., Gehrke J, 2000"Querying the physical world," *IEEE Personal Communication*.
- [6] C. Park and P. H. Chou. Ambimax: Autonomous energy harvesting platform for multisupply wireless sensor nodes. In Third Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks, pages 168-177, VA Reston, USA, September 2006.
- [7] J. Hsu, S. Zahedi, A. Kansal, M. Srivastava, and V. Raghunathan: Adaptive duty cycling for energy harvesting systems. In International Symposium on Low Power Electronics and Design, pages 180-185, Tegernsee, Germany, October 2006.
- [8] Luca P. Carloni, Fernando De Bernardinis, Alberto L. Sangiovanni-Vincentelli, and Marco Sgroi. The art and science of integrated systems design. In Proceedings of the 28th European Solid-State Circuits Conference, ESSCIRC 2002, pages 19-30, Firenze, Italy, September 2002.
- [9] Akkaya K. and Younis M., 2005"A survey on routing protocols for wireless Sensor network," *journal of Adhoc Networks*, vol 3, 325-349 .

- [10] Heinzelman W.R., Chandrakasan A, and Balakrishnan H., 2000"Energy Efficient Communication Protocol for Wireless Micro sensor Networks," *Proc. 33rd Hawaii Int'l. Conf. Sys. Sci.*
- [11] Lin SHEN and Xiangquan SHI: A Location Based Clustering Algorithm for Wireless Sensor Networks, INTERNATIONAL JOURNAL OF INTELLIGENT CONTROL AND SYSTEMS,VOL. 13, NO. 3, SEPTEMBER 2008,208-213
- [12] Seapahn Megerian and Miodrag Potkonjak, "Wireless sensor networks," Book Chapter in Wiley Encyclopedia of Telecommunications, Editor: John G. Proakis, 2002.
- [13] OMNET ++ Website, www.omnetpp.org.
- [14] M. Bani Yassein, A. Al-zou'bi,Y.Khamayesh,W. Mardini: "Improvement on LEACH Protocol of Wireless Sensor Network(VLEACH)."
- [15] F. Simjee and P. H. Chou. Everlast: Long-life, supercapacitor-operated wireless sensor node. In International Symposium on Low Power Electronics and Design, pages 197-202,Tegernsee, Germany, October 2006
- [16] P. Zhang, C. M. Sadler, S. A. Lyon, and M. Martonosi. Hardware design experience in zebanet. In Second International Conference on Embedded Networked Sensor Systems, pages 227-238, Baltimore, USA, 2004.
- [17] V. Loscri, G. Morabito and S. Marano. "A Two-Levels Hierarchy for Low-Energy Adaptive Clustering Hierarchy".
- [18] J.E.Wieselthier, G.D. Nguyen and A. Ephremides, On the construction of energy-efficient broadcast and multicast trees in wireless networks,2000.
- [19] Mudasser Iqbal , Iqbal Gondal, Laurence Dooley: "An Energy-Aware Dynamic Clustering Algorithm for Load Balancing in Wireless Sensor Networks" 2006.
- [20] Sajid Hussain and Abdul W. Matin : "Hierarchical Cluster-based Routing in Wireless Sensor Networks."
- [21] W. Heinzelman, "Application-Specific Protocol Architectures for Wireless Networks," PhD Thesis, MIT, June 2000
- [22] Jun Wang , Yong-Tao Cao , Jun-Yuan Xie , Shi-Fu Chen : "Energy Efficient Backoff Hierarchical Clustering Algorithms for Multi-Hop Wireless Sensor Networks", JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 2011.

AUTHORS PROFILE



Ashim Kumar Ghosh received B.Sc. in engineering degree in Computer Science and Engineering (CSE) from Dhaka University of Engineering and Technology (DUET), Gazipur-1700, Bangladesh. His research interest includes Communication protocols for wireless sensor networks, Mobile computing, Intelligent transportation system, and Parallel & distributed computing. Contact him at ashim.cse06@gmail.com.



Anupam Kumar bairagi obtained B.Sc. degree in Computer Science and Engineering (CSE) from Khulna University (KU), Khulna-9208, Bangladesh. Now he is a faculty member of Department of CSE, KU. His key research interest in Cryptography, Networks and Web Security, Communication protocols for wireless sensor networks, Mobile computing, and Intelligent transportation system. E-mail at cse9620@gmail.com.



Dr. M. Abul Kashem has been serving as an Associate Professor and Head of Department of Computer Science and Engineering (CSE), Dhaka University of Engineering and Technology (DUET) Gazipur, Bangladesh. His key research interest in Speech Signal Processing, Communication protocols for wireless sensor networks and Intelligent transportation system. E-mail at drkashemll@duet.ac.bd



Md. Rezwan-ul-Islam received B.Sc. in engineering degree in Computer Science and Engineering (CSE) from Dhaka University of Engineering and Technology (DUET), Gazipur-1700, Bangladesh. His research interest Wireless sensor networks, Artificial intelligent, Expert System, Intelligent transportation system and Cryptography. Contact him at shuvo.rezwan@gmail.com.



A J M Asraf Uddin received B.Sc. in engineering degree in Computer Science and Engineering (CSE) from Dhaka University of Engineering and Technology (DUET), Gazipur-1700, Bangladesh. His research interest Wireless sensor networks, Web security, Heterogeneous network with mobility. Contact him at asrafce08@gmail.com.

Eyes Based Electric Wheel Chair Control System

- I (eye) can control Electric Wheel Chair -

Kohei Arai

Department of Information Science
Saga University
Saga City, Japan

Ronny Mardiyanto

Graduate School of Science and Engineering
Saga University
Saga City, Japan

Abstract— Eyes base Electric Wheel Chair Control: EBEWC is proposed. The proposed EBEWC is controlled by human eyes only. Therefore disabled person can control the EBEWC by themselves. Most of the computer input system with human eyes only consider in specific condition and does not work in a real time basis. Moreover, it is not robust against various user races, illumination conditions, EWC vibration, and user's movement. Though experiments, it is found that the proposed EBEWC is robust against the aforementioned influencing factors. Moreover, it is confirmed that the proposed EBEWC can be controlled by human eyes only accurately and safely.

Keywords-computer input by human eyes only; gaze estimation; electric wheelchair control.

I. INTRODUCTION

The existing computer input devices such as keyboard, mouse, and the other input devices have been used to interact with digital instruments. These computer input devices cannot be operated by handicap persons. In this paper, a computer input device by human eyes only is proposed for handicap person and also for wearable computing.

The existing computer input devices without finger, voice, and gesture can be divided into five categories;

- (1) Bio-potential based method which utilizes potential from user's body actions acquired by using special instrument. Instrument such as Electrooculograph (EOG), Electromyograph (EMG) [1], and Electroencephalograph (EEG) [2], Search coil can be used for measuring bio-potential. The search coil output can be used as sources of computer input for handicap person. EOG method [7], [16] uses voltage differences between fore and aft surface of eyes.
- (2) Voice Based method [3], which use user's voice as source input. Voice analysis is used to analyze user's voice and convert into digital data. The weakness of this system is vulnerable against noise. Other voices which come from surrounding user may affect the system.
- (3) Motion based method [4], utilizes other normal movement organs to operate computer input. Head, foot, and etc can be used to control computer input.
- (4) Image Analysis method [10]-[15], utilizes camera to analyze user's desire and convert into digital data.

Several image processing methods are used to analyze user's desire. The user's desire itself can be done by Gaze based [5], [6], [9], analyze user's desire from users gaze, Face based analyze user's desire from face expression, and the others.

- (5) Search coil method [8] uses induced voltage with coil including in contact lenses attached to user's eyes.

Methods (1) and (5) insists psychological and physical burden to users because these methods require direct sensors which are attached to user's face. Also these methods are relatively costly. On the other hand, the image analysis method does not insist any load to users and is realized with comparatively cheap cost. For the aforementioned reason, we propose an image analysis based method.

Electric Wheel Chair: EWC for assisted mobility is presented [16]-[20]. There is the proposed system for assisted mobility using eye movements based on Electrooculography. EWC is controlled by eye movement which is acquired using Electrooculograph [16]. Also there is the proposed integrated solution to motion planning and control with input from three sources [17]. At the highest level, the human operator can specify goal destination using visual interface by pointing to location or features on display. This input lets the chair automatically generate a deliberative plan incorporating prior knowledge. At the intermediate level, the human operator must use reactive controller to avoid obstacle and features that the sensor detect. At the lowest level, the human operator can directly provide velocity command using joystick. There is the proposed vision-based navigation for an electric wheelchair using ceiling light landmark [18]. The wheelchair is equipped with two cameras those are used for self-location and obstacle avoidance. The fluorescent ceiling lights are chosen as landmarks since they can be easily detected and do not require an additional installation. Also there is the proposed head gesture based control of an intelligent wheelchair [19]. This system used Adaboost face detection. By detecting frontal face and nose position, head gesture is estimated and used to control the wheelchair.

There is the proposed EWC control with gaze direction and eye blinking [20]. The gaze direction is expressed by horizontal angle of gaze, and it is derived from the triangle form formed by the center position of eyes and nose. The gaze direction and eye blinking are used to provide the direction and timing

command. The direction command related to the movement direction of EWC and the timing command related to the time condition when the EWC should move.

In this paper, we propose computer input system with human eye-only and used for controlling EWC. It is called Eye Based EWC: EBEWC hereafter. EBEWC works based on eye gaze. When user looks at appropriate angle/key, then computer input system will send command to EWC. EBEWC is controlled by three keys: left (Turn left), right (Turn right), and down (Move forward) in order to give command to electric wheelchair: turn left, turn right, and go forward. This three combination keys are more safely than others combination keys. The stop key is not required because EWC will automatically stop when user change the gaze. Gaze estimation system still faced with problem such as robustness against a variety of user types, accuracy, illumination changes, vibration, and calibration.

In order to estimate gaze based on image analysis, it is common that gaze location is estimated with pupil location. The previously published pupil detection methods can be divided into two categories: the active Infrared: IR-based approaches [22], [23], [31] and the traditional image-based passive approaches [24]-[30]. Eye detection based on Hough transform is proposed [24]-[26]. Hough transform is used for finding the pupil. Eye detection based on motion analysis is proposed [21], [22]. Infrared lighting is used to capture the physiological properties of eyes (physical properties of pupils along with their dynamics and appearance to extract regions with eyes). Motion analysis such as Kalman filter and mean shift which are combined with Support Vector Machine: SVM used to estimate pupil location. Eye detection using adaptive threshold and morphologic filter is proposed [27]. Morphologic filter is used to eliminate undesired candidates for an eye. Hybrid eye detection using combination between color, edge and illumination is proposed [30].

In order to estimate the gaze, we use pupil knowledge for improvement on robustness against different users. In gaze estimation based on image analysis, almost all utilize pupil location as reference. In this stage, pupil detection accuracy is very important because all the gaze calculations are made based on the detected pupil location. Many researches did not give much attention on this part because most of them use ideal images as the source in order to find pupil location. When user looks at forward, pupil can be detected clearly. Meanwhile, it is usually not so easy to detect pupil location when user looks the other directions. A variety of skin colors, races, interference with eye lids, disappearances, and changes of eye shape (when eye move) make the pupil detection very difficult. Pupil knowledge such as shape, size, location, and motion are used in the proposed method. This knowledge works based on the knowledge priority.

Pupil appearance such as size, color, and shape are used as first priority. When this step fails, then pupil is estimated based on its location as second priority. When all steps fails, pupil is estimated based on its motion as last priority. Last, we use eye model in order to convert pupil location into gaze direction. This gaze direction will be used to control EWC. In order to

improve the robustness against user's movement, illumination changing, and vibration, IR camera mounted glass is used.

The proposed system is tested by several users with different race and nationality. The experimental results with the proposed eye gaze estimation method are compared to well-known adaptive threshold method and template matching method. Also robustness against illumination changing, noise influence, vibration, and accuracy has been confirmed. In the section 2, the proposed EBEWC system is described followed by some experimental results. Then section 4 describes some discussions followed by conclusion.

II. PROPOSED METHOD

The problem of the utmost importance of a proposed EBEWC system is the robustness against different user types, illumination changes, user's movement, vibration, and accuracy. In order to consider these as vehicle system, if the user changes, the system should be works without any input parameter changes. In accordance with EWC movement, illumination condition may change. Also, disturbances due to EWC vibration is potentially problem.

In the conventional EWC control system with human eyes only, camera is mounted on EWC. This may cause a vulnerable when EWC is vibrated. Also, when user moves their head, gaze estimation is difficult. Furthermore, illumination condition may change during EWC movement. The proposed EBEWC system utilizes IR camera which mounted on user's glass. This way will eliminate problems of illumination changes, user's movement, and EWC vibration. Furthermore, the pupil detection based on pupil knowledge will improve the robustness against different users.

A. Hardware Configuration

Hardware configuration of the proposed EBEWC is shown in Fig.1.

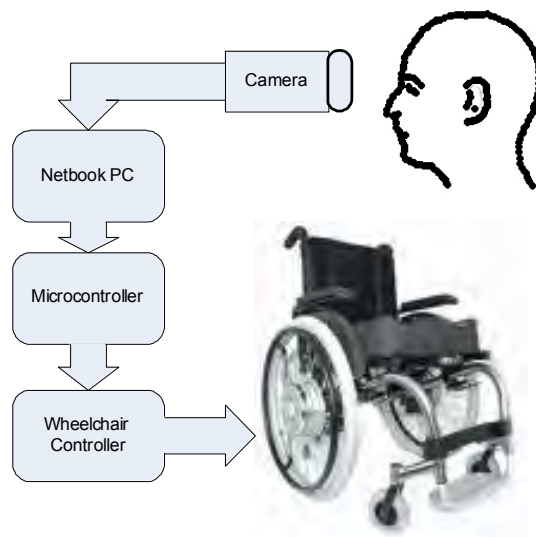


Figure 1. Hardware Configuration

Netbook of Asus EeePC with 1.6 GHz Intel Atom processor, 1GB of RAM, 160GB hard drive, and run Windows

XP Home edition is used. We develop our software under C++ Visual Studio 2005 and Intel provided OpenCv image processing library [21]. The proposed EBEWC system utilizes infrared web camera, NetCowBoy DC-NCR 131 as face image acquisition in a real time basis. This camera has IR Light Emission Diode: LED. Therefore it is robust against illumination changes. In order to allow user movement and EWC vibration, IR Camera mounted glass is used. The distance between camera and eye is set at 15.5 cm as shown in Fig.2. The EBEWC uses Yamaha JWX-1 type of EWC which is controlled by human eyes only through microcontroller of AT89S51. This microcontroller can convert serial output from Netbook to digital output for control.

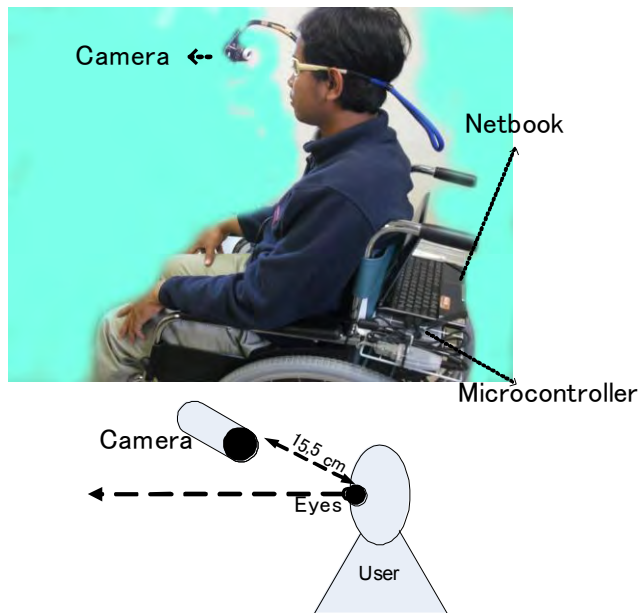


Figure 2. Electric Wheelchair Hardware

In order to control EWC using Netbook PC, custom micro controller circuit is used to modify standard control of Yamaha EWC. Default control of the EWC is made by joystick. Micro controller is replaced to the joystick. Command is delivered by Netbook PC, and then micro controller works to move EWC. USB interface on Netbook PC is used to connect with the other peripheral. The interface of the micro controller circuit is RS232. To connect between Netbook PC and the micro controller, USB to Serial converter is used. The micro controller is driven by the relay equipped with the EWC. Micro-controller connection is shown in Fig.3.

B. Gaze Estimation

In order to estimate gaze, eye should be detected and tracked. Fig. 4 shows the process flow of eye detection and tracking. The proposed EBEWC system detect eye based on deformable template method [32]. This method matches between eye template and source images. We create eye template and apply Gaussian smoother. Deformable template method detects rough position of eye. Benefit of deformable template method is that it takes less time than classifier methods. Although this method is faster than the other classifier methods, the aforementioned robustness is not good enough.

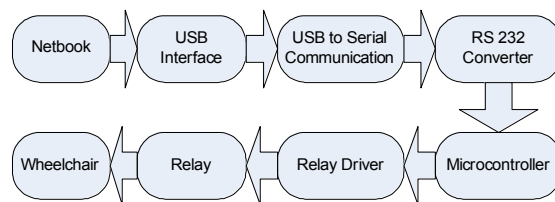


Figure 3 Microcontroller AT89S51 connects to other peripheral through serial communication. Serial communication type should be converted to USB communication using USB to serial converter

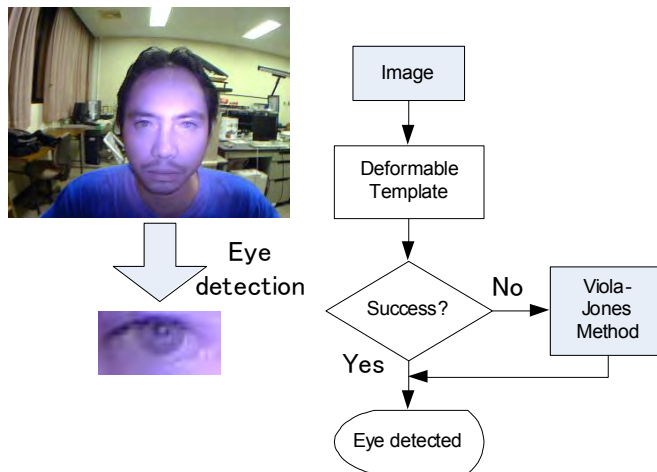


Figure 4. Flow of Eye Detection, eye is detected by using combination between deformable template and Viola-Jones methods.

In the proposed EBEWC system, the well known Viola-Jones classifier in the OpenCV library [21] detects eye when the deformable template fails to detect eye position. The Viola-Jones classifier employs Adaboost at each node in the cascade to learn a high detection rate the cost of low rejection rate multi-tree classifier at each node of the cascade. The Viola-Jones function in OpenCV is used for the proposed EBEWC system. Before using the function, we should create XML file through learning processes. The training sample data (face or eye image) must be collected. There are two sample types,; negative and positive samples. Negative sample corresponds to non-object images while positive sample corresponds to object image. After acquisition of image, OpenCV will search the face center location followed by search the eye center. By using combination between deformable eye template and the Viola-Jones method, eye location will be detected. Advantages of this proposed method is fast and robust against circumstances changes.

After the roughly eye position is founded, eye image is locked and tracked. Therefore, there is no need to detect eye any more. The detected eye position is not going to be changed because the camera is mounted on the glass. Eye gaze is estimated based on pupil center location. Because of this system rely on the pupil center location; pupil detection has to be done perfectly. Pupil is detected by using its knowledge. Process flow of the pupil detection is shown in Fig.5. Three types of knowledge are used. We use pupil size, shape, and color as the first knowledge. First, adaptive threshold method is applied for pupil detection. Threshold value T is determined by the average pixel value (mean) of eye image μ . We set the

threshold value of 27% bellow from the average value empirically.

$$\mu = \frac{1}{N} \sum_{i=0}^{N-1} I_i \quad (1)$$

$$(2)$$

$$T = 0.27\mu$$

Pupil is signed as one black circle in the acquired eye image. In the first stage, pupil is detected with adaptive threshold method when the pupil is clearly appears in the acquired eye image. By using connected labeling component, we can easily estimate the pupil center location. Meanwhile, noise is usually appears in the image, we can distinguish them by estimate its size and shape as knowledge. The situation is shown in Fig.6 (a). In another case, shape and size of eye is changed when user looks to right and left directions. In this situation, pupil detection is hard to find. Noise and interference due to eyelid have to be eliminated. In this situation, knowledge of size and shape can be used. Also previously detected pupil center can be used when user close eye. The pupil center location is determined using the following equation,

$$P(t-1) - C < P(t) < P(t-1) + C \quad (3)$$

The reasonable pupil location $P(t)$ is always in surrounding previous location $P(t-1)$ with the area C . Fig.6 (b) shows such this situation.

When the entire above step is fail to detect pupil center location, and then we estimate pupil center location by using eye motion. This situation is happened when the black pixels are mixed with others or no black pixel at all in the acquired image. We employ this knowledge as a last priority to avoid an ambiguous motion that causes misidentification of pupil to the other eye components. We tracked a pupil location using its previous location based on the well-known Kalman filter [23],[33]. Kalman Filter corrects the estimated pupil location and velocity. In each pixel as detected as a pupil, we assume the location of pupil is (i_b, j_t) and velocity is (u_b, v_t) . Because of the previous location is (i_b, j_t) , the current location should be $(i_t + u_b, j_t + v_t)$. We can model the state of pupil location is as follows,

$$x_t(i, j) = Ax_{t-1}(i, j) + \omega_{k-1}(i, j) \quad (4)$$

where, x_t is actual state and A is state transition while ω_k denotes additive noise. Next, we assume that the estimated location is (\hat{i}_t, \hat{j}_t) . The measurement process can be modeled as follows,

$$z_k(i, j) = Hx_k(i, j) + v_k(i, j) \quad (5)$$

where v_k represents noises in the measurement, H represents observation matrix. This method works when the estimated pupil location becomes entrusted. Such a condition may happen when the other components disturb the pupil location estimation. By using time updating algorithm and

measurement update process, a better estimated pupil location will be obtained.

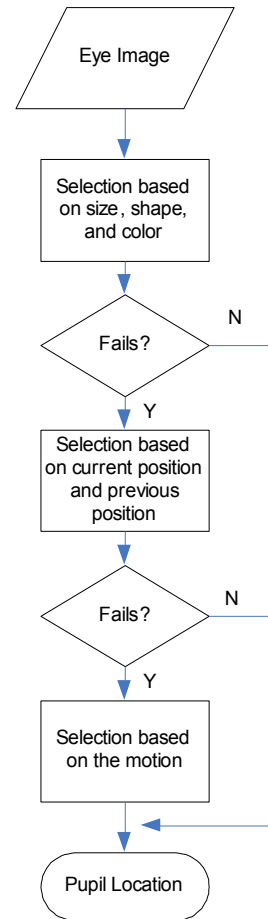
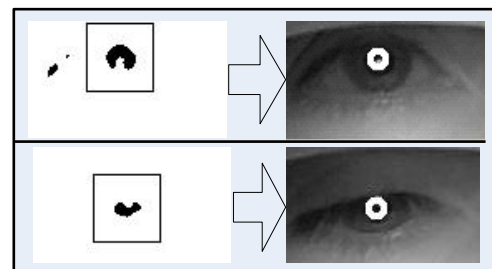
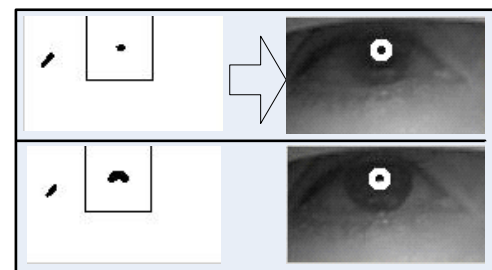


Figure 5. Flow of pupil detection



(a) Case 1, Pupil clearly appears



(b) Case 2, Pupil clearly appears with some defect by eyelid

Figure 6 Examples of the acquired pupil images

C. Eye Model

A simple eye model is defined as shown in Fig.7.

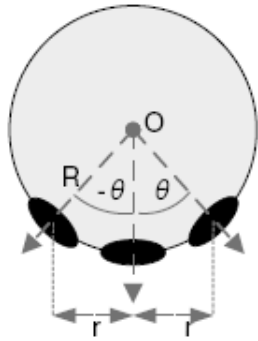


Figure 7. Shape model of eye

The eyeball is assumed to be a sphere with radius R . Although the typical eyeball shape is ellipsoid, sphere shape assumption does not affect to the pupil center location estimation so much that spheroid shape of assumption of eyes ball shape does not affect too much to the pupil center location estimation accuracy. The pupil is located at the front of eyeball. The distance from the center gaze to current gaze is assumed to be r . Gaze is defined as angle θ between normal gaze and r . The relation between R , r and θ is as follows,

$$r = R \sin \theta \tag{6}$$

$$\theta = \arcsin\left(\frac{r}{R}\right) \tag{7}$$

The typical radius of the eyeball ranges from 12 mm to 13 mm according to the anthropometric data [34]. Hence, we use the anatomical average assumed [35] into the proposed algorithm. Once r has been found, gaze angle θ is easily calculated. In order to measure r , the normal gaze is should be defined. In the proposed EBWC system, when the system runs as the first time, the user has to look at the computer display then user looks at the center of the computer screen. At this moment, we record that this pupil location is normal gaze position. In order to avoid error when acquire normal gaze, it is verified by compare between its value and center of two eye corners. Next if user look at same key within 0.7 second, then gaze value will be send to EWC. This way will avoid noise gaze which is caused by user intention is not always focus in same key.

D. EWC Control

In order to control EWC, we design three functional keys invisible layout, move forward, turn right, and turn left at the three specific portions. If user looks at the other portion, then EWC is stopped for safety reason. There is no need to display the key layout at all. Users understand the location of desired key so that it can be selected. For instance, when user looks at the right direction for more than 0.7 second, EWC is then turn right until user changes the gaze direction. If user keep look at the same direction, then EWC is continued the previously determined moving action. Fig.8 shows the proposed key layout.

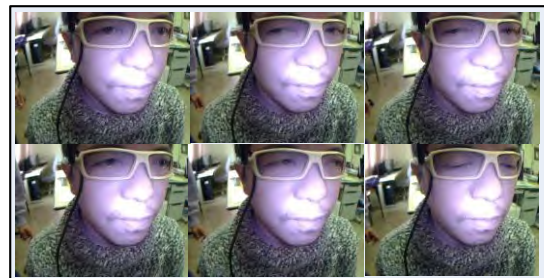
Stop	Stop	Stop
Turn left	Forward	Turn right
Stop	Stop	Stop

Figure 8. 3 by 3 of key layout for EWC control

III. EXPERIEMNTS

A. Exampel of Acquired Face Images

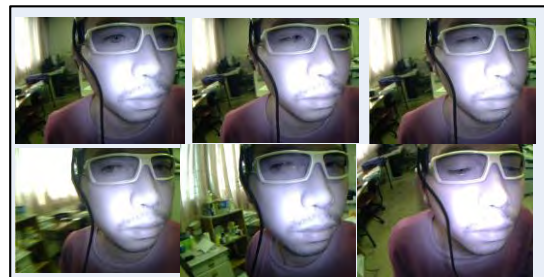
Fig.9 shows the examples of the acquired face images of the three different Indonesian users. The experiments are carried out with six different users who have the different race and nationality: Indonesian, Japanese, Sri Lankan, and Vietnamese. We collect data from each user during user makes several eye movements and EWC actions.



(a)Indonesian No.1



(b)Indonesian No.2

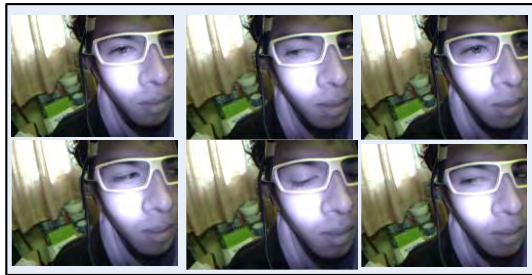


(c)Indonesian No.3

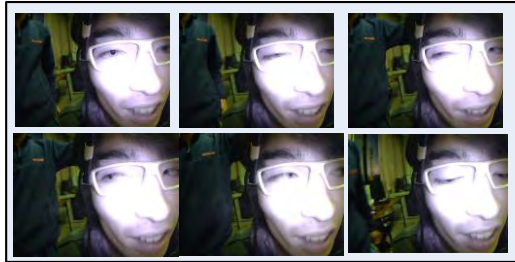
Figure 9. Example of Indonesian Images

The collected data contain several eye movement such as look at forward, right, left, down, and up. Two of Indonesians have width eye and clear pupil. The number of images is 552 samples and 668 samples. Another Indonesian has slanted eyes (Off-axis viewing) and the pupil is not so clear. The number of images is 882 samples. We also collected the data for Sri Lankan people as shown in Fig.10 (a). His skin color is black with thick eyelid. The number of images is 828 samples. The collected data of Japanese is shown in Fig.10 (b). His skin

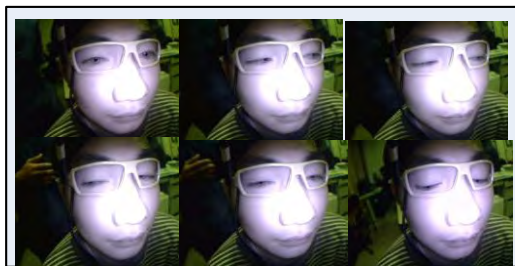
color is bright with slanted eyes. The number of images is 665 samples. The Vietnamese data is shown in Fig.10 (c).



(a) Srilankan Images



(b) Japanese Images



(c) Vietnamese Images

Figure 10 Examples of the acquired face images.

B. Success Rate of Pupil Detection (Gaze Estimation)

Performance of success rate of pupil detection and gaze estimation is evaluated with six users. The success rate of the proposed EBWC method is compared to the conventional adaptive threshold method and template matching method. The adaptive threshold method is modified with inclusion of a connected labeling method. The template matching method uses pupil template as a reference and allows matching with the currently acquired images. The results of success rate evaluation experiments are shown in Table 1. The proposed EBWC method is superior to the other conventional methods. Also it is confirmed that the EBWC method is robust against various user types with the variance of 16.27 as shown in Table 1.

C. Influence Due to Illumination Changes

The aforementioned success rate is evaluated with changing the illumination conditions. The experimental results are shown in Fig.11. As shown in Fig.11, 0-1200 LUX of illumination changes do not affect to the success rate while much greater than 1600 LUX of illumination condition may affect to the success rate. Because IR Light Emission Diode: LED utilized IR camera is used for the proposed EBWC system, low

illumination is appropriate rather than high illumination. As a result, it is found that the proposed EBWC system does work in the normal illumination condition ranges from 0 to 1500 LUX.

TABLE 1. ROBUSTNESS AGAINST VARIOUS USERS, THIS TABLE SHOWS THAT OUR METHOD ROBUST ENOUGH AGAINST VARIES USER AND ALSO HAS HIGH SUCCESS RATE

User Type	Nationality	Adaptive Threshold (%)	Template Matching (%)	Proposed Method (%)
1	Indonesian	99.85	63.04	99.99
2	Indonesian	80.24	76.95	96.41
3	Srilankan	87.80	52.17	96.01
4	Indonesian	96.26	74.49	99.77
5	Japanese	83.49	89.10	89.25
6	Vietnamese	98.77	64.74	98.95
Average		91.07	70.08	96.73
Variance		69.75	165.38	16.27

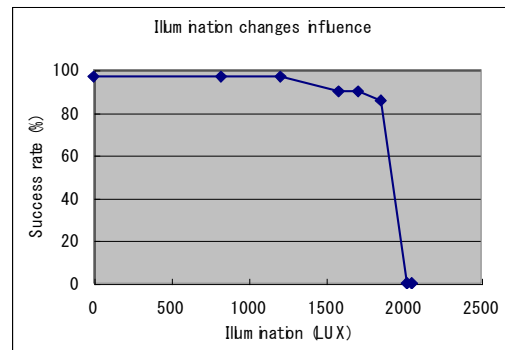


Figure 11 Influence against illumination changes

D. Influence Due to EWC Vibrations

Because EWC vibrates usually, influence due to vibration on the success rate and the acceleration to user has to be clarified Four types of vibrations are taken into account. Shock of acceleration in unit of m/s^2 in direction of x , y , and z at the HMD and IR camera mounted glass is measured with acceleration measurement instruments. x direction corresponds to the forward direction while y direction corresponds to the side direction as well as z direction corresponds to the up and down direction. The experimental results are shown in Fig.12. During user controls EWC with the proposed EBWC system, EWC moves 10 m of collider. On the collider, there are some steps and it may cause the vibrations. x direction of acceleration is always $10 m/s^2$ because the EWC is moving forward. The measured accelerations for the first two vibrations are $25 m/s^2$ while the third acceleration for the third vibration is $35 m/s^2$ and the fourth acceleration for the fourth vibration is $40 m/s^2$.

Even vibration/ shock are happened; user body absorbed the vibrations. Furthermore, user wears the HMD and camera mounted glass so that the relative locations among HMD and eye as well as pupil are stable enough. Therefore, success rate of pupil detection is not changed results in no influence due to vibration on gaze estimation accuracy.

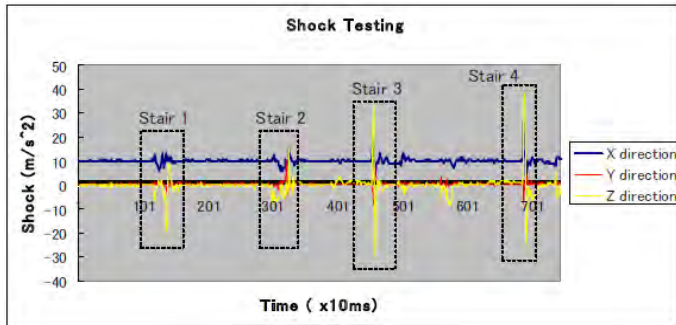


Figure 12. Shock Influence

E. Process Time Required for Eye Detection and Gaze Estimation

It is confirmed that the proposed EBECW does work in a real time basis. The time required for eye detection is measured. It is 286.92 ms while that for gaze estimation is 14.08 ms

F. Influence Due to Noise on Success Rate

Influence due to noise on success rate is evaluated with random number generator of Mesenne Twister. Normal distribution with zero mean and 0-70 standard deviation of random number is used for the experiment. Success rate of pupil detection and gaze estimation is shown in Fig.13. As the result of the experiment, it is found that the proposed EBECW system has 10 % degradation of success rate. Therefore, user has to input gaze twice for 1 time out of 10 times.

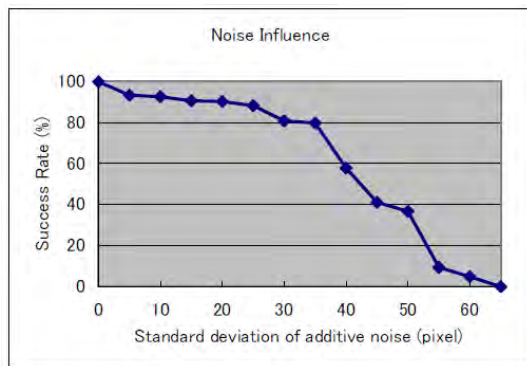


Figure 13. Influence due to noise on success rate

G. Gaze Estimation Accuracy

Gaze estimation accuracy is measured at the middle center to horizontally aligned five different locations with the different angles as shown in Fig.14. The experimental result shows the gaze estimation error at the center middle (No.1) shows zero while those for the angle ranges from -5 to 15 degrees is within 0.2 degree of angle estimation error. Because the user looks at the center middle with 5 degree allowance

results in 0.2 degree of gaze estimation accuracy. Fig.15 shows the outlook of the proposed EBECW system.

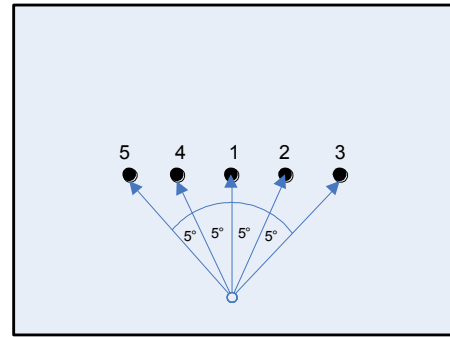


Figure 14. Measuring accuracy

TABLE 2 GAZE ESTIMATION ACCURACY AT THE DESIGNATED VIEWING ANGLES

Point	1	2	3	4	5
Error(degree)	0	0.2	0.2	0.2	3.12



Figure 15. Testing of EWC

IV. CONCLUSION

Specific features of the proposed EBECW system are,

- (1) It allows user movement: User can move during using the proposed EBECW system in directions in the allowable distance of which the camera mounted on the glass acquires user face,

- (2) It does not require any calibration process before using the proposed EBEWC system,
- (3) It is robust against immunization changes, additive noises, vibrations to the EWC, user nationality (different color of eye as well as skin, size, and shape of eye), eyelid influence, and shade and shadow,
- (4) Pupil center detection accuracy for the acceptable angle range is almost 0.2 degree,
- (5) Even if pupil center is not detected when user close eye, gaze location is estimated with the knowledge of previously detected pupil center location.

In the near future, we will conduct fatigue measurement experiments with the proposed EWC control by human eyes only. Then we can ensure using the EWC for a long time period.

REFERENCES

- [1] Microsoft Research patents, <http://www.engadget.com/2010/01/03/microsoft-research-patents-controller-free-computer-input-via-em/>.
- [2] Culpepper, B.J, Keller, R.M: "Enabling computer decisions based on EEG input", IEEE Trans on Neural Systems and Rehabilitation Engineering, 11, 354-360, [2003].
- [3] United States Patent, <http://www.freepatentsonline.com/4677569.html>.
- [4] David L. Jaffe: "An ultrasonic head position interface for wheelchair control", Journal of Medical Systems, 6, 4, 337-342, [1982].
- [5] K. Abe, S. Ohiamd M. Ohyama: "An Eye-gaze Input System based on the Limbus Tracking Method by Image Analysis for Seriously Physically Handicapped People", Proceedings of the 7th ERCIM Workshop "User Interface for All" Adjunct Proc., 185-186, [2002].
- [6] <http://www.creact.co.jp/jpn/por.pdf>
- [7] Kuno, Yagi, Fujii, Koga, Uchikawa "Development of the look input interface using EOG", the Information Processing Society of Japan paper magazine C39C5, 1455-1462, [1998].
- [8] D.A. Robinson, "A method of measuring eye movement using a sclera search coil in a magnetic field", IEEE Trans. on Biomedical Electronics, 10, 137-145, [1963].
- [9] <http://webvision.med.utah.edu/>
- [10] Ito, Nara: "Eye movement measurement by picture taking in and processing via a video capture card, an Institute of Electronics", Information and Communication Engineers Technical Report, 102, 128, 31-36, [2002].
- [11] Kishimoto, Yonemura, Hirose, Changchiang: "Development of the look input system by a cursor move system", Letter of the Institute of Image Information and Television Engineers, 55, 6, 917-919, [2001].
- [12] Corno, L.Farinetti, I. Signorile.: "A Cost-Effective Solution for Eye-Gaze Assistive Technology", Proceedings of the IEEE International Conf. on Multimedia and Expo, 2, 433-436, [2002].
- [13] Abe, Ochi, Oi, Daisen: "The look input system using the sclera reflection method by image analysis", Letter of the Institute of Image Information and Television Engineers, 57, 10, 1354-1360, [2003].
- [14] Abe, Daisen, Oi: "The multi-index look input system which used the image analysis under available light", Letter of the Institute of Image Information and Television Engineers, 58, 11, 1656- 1664, [2004].
- [15] Abe, Daisen, Oi: "The look input platform for serious physically handicapped persons", Human Interface Society Human interface symposium 2004 collected papers, 1145-1148, [2004].
- [16] Barea, R., Boquete, L., Mazo, M., Lopez, E.: "System for Assisted Mobility using Eye Movements based on Electrooculography", IEEE Transaction on Neural System and Rehabilitation Engineering, 10, 4, 209-218, [2002].
- [17] Sarangi, P., Grassi, V., Kumar, V., Okamoto, J.: "Integrating Human Input with autonomous behaviors on an Intelligent Wheelchair Platform", Journal of IEEE Intelligent System, 22, 2, 33-41, [2007].
- [18] Wang, H., Ishimatsu, T.: "Vision-based Navigation for an Electric Wheelchair Using Ceiling Light Landmark", Journal of Intelligent and Robotic Systems, 41, 4, 283-314, [2005].
- [19] P. Jia and H. Hu: "Head Gesture based control of an Intelligent Wheelchair", Proceedings of the 11th Annual Conference of the Chinese Automation and Computing Society in the UK [CACSUK05], 85-90, [2005].
- [20] D. Purwanto, R. Mardiyanto, K. Arai: "Electric wheelchair control with gaze direction and eye blinking", Proceedings of The Fourteenth International Symposium on Artificial Life and Robotics, GS21-5, B-Con Plaza, Beppu, [2008].
- [21] Gary Bradski, Andrian Kaebler: "Learning Computer Vision with the OpenCV Library", O'REILLY, 214-219, [2008].
- [22] Haro. A, Flickner. M, Essa, I: "Detecting and Tracking Eyes By Using Their Physiological Properties, Dynamics, and Appearance ", Proceedings of the CVPR 2000, 163-168, [2000]
- [23] Zhiwei Zhu, Qiang Ji, Fujimura, K., Kuangchih Lee: "Combining Kalman filtering and mean shift for real time eye tracking under active IR illumination", Proceedings of the 16th Pattern Recognition International Conference, 4, 318- 321, [2002].
- [24] Takegami. T, Gotoh. T, Kagei. S, Minamikawa-Tachino. R: "A Hough Based Eye Direction Detection Algorithm without On-site Calibration", Proceedings of the 7th Digital Image Computing: Techniques and Applications, 459-468, [2003].
- [25] K.M.Lam, H.Yan, Locating and extracting eye in human face images, Pattern Recognition, 29, [5], 771-779, [1996].
- [26] G.Chow, X.Li, Towards a system of automatic facial feature detection, Pattern Recognition 26, 1739-1755, [1993].
- [27] Rajpathaka. T, Kumarb. R, Schwartzb. E: Eye Detection Using Morphological and Color Image Processing", Proceedings of the Florida Conference on Recent Advances in Robotics, [2009].
- [28] R.Brunelli, T.Poggio, Face Recognition, Features versus templates, IEEE Trans. Patt. Anal. Mach. Intell. 15, 10, 1042-1052, [1993].
- [29] D.J.Beymer, Face Recognition under varying pose, IEEE Proceedings of the Int. Conference on Computer Vision and Pattern Recognition [CVPR'94], Seattle, Washington, USA, 756- 761, [1994]
- [30] Shafi. M, Chung. P. W. H: "A Hybrid Method for Eyes Detection in Facial Images", International Journal of Electrical, Computer, and Systems Engineering, 231-236, [2009].
- [31] Morimoto, C., Koons, D., Amir, A., Flickner. M: "Pupil detection and tracking using multiple light sources", Image and Vision Computing, 18,4, 331-335, [2000].
- [32] A. Yuille, P. Haallinan, D.S. Cohen: "Feature extraction from faces using deformable templates", Proceedings of the IEEE Computer Vision and Pattern Recognition, 104-109, [1989].
- [33] <http://citeseer.ist.psu.edu/443226.html>
- [34] K.-N. Kim and R. S. Ramakrishna: "Vision-based eye gaze tracking for human computer interface", Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, 2, 324-329, [1999].
- [35] R. Newman, Y. Matsumoto, S. Rougeaux and A. Zelinsky: "Real-time stereo tracking for head pose and gaze estimation", Proceedings of the Fourth International Conference on Automatic Face and Gesture Recognition, 122-128, [2000].

AUTHORS PROFILE

Kohei Arai received a PhD from Nihon University in 1982. He was subsequently appointed to the University of Tokyo, CCRS, and the Japan Aerospace Exploration Agency. He was appointed professor at Saga University in 1990. He is also an adjunct professor at the University of Arizona since 1998 and is Vice Chairman of ICSU/COSPAR Commission A since 2008.

Fuzzy Petri Nets for Human Behavior Verification and Validation

M. Kouzehgar, M. A. Badamchizadeh, S. Khanmohammadi
Department of Control Engineering
Faculty of Electrical and Computer Engineering, University of Tabriz,
Tabriz, Iran

Abstract— Regarding the rapid growth of the size and complexity of simulation applications, designing applicable and affordable verification and validation (V&V) structures is an important problem. On the other hand, nowadays human behavior models are principles to make decision in many simulations and in order to have valid decisions based on a reliable human decision model, first the model must pass the validation and verification criteria. Usually human behavior models are represented as fuzzy rule bases. In all the recent works, V&V process is applied on a ready given rule-base. In this work, we are first supposed to construct a fuzzy rule-base and then apply the V&V process on it. Considering the professor-student interaction as the case-study, in order to construct the rule base, a questionnaire is designed in a special way to be transformed to a hierarchical fuzzy rule-base. The constructed fuzzy rule base is then mapped to a fuzzy Petri net and then within the verification (generating and searching the reachability graph) and validation (reasoning the Petri net) process is searched for probable structural and semantic errors.

Keywords- human behavior; verification; validation; high-level fuzzy Petri nets; fuzzy rules.

I. INTRODUCTION

All On the whole nowadays human behavior models are principles to make decision in many simulations. In order to improve the fidelity and automation of simulation exercises, human behavior models have become key components in most simulations. Our work firstly suggests the setup of the students' deduction system and looking at the student-professor interaction as a system; furthermore the application of verification and validation in education system is tested. The professor-student interaction is chosen as the case-study since it is something tangible for all of us who really live in an academic environment being in direct contact with such a system in our everyday life and something you must deal with a day after the other really was worth giving a try.

It's high time simulating human behavior and presenting relevant control schemes has become an exciting field for the researchers. Human behavior modeling or human behavior representation (HBR) is a field of study important in military service research [1, 2], robotics [3], brain-computer interface (BCI) , human machine interface (HMI) [4, 5] and some specially oriented anthropology studies [6]. Human behavior models are often represented by finite state machines, rules, fuzzy rules [7], artificial neural networks [8], fuzzy hybrid rule-frames [9], fuzzy dynamic Bayesian networks [10, 11],

concept lattice [12], multi-agent based modeling [13]. Among all these, HBR by a fuzzy rule base is the most common [14].

In order to have a reliable human behavior model on which many decisions depends, it is essential to ensure that the HBM passes through V&V criteria. In this research V&V for a fuzzy rule-base is focused. Some techniques for verification of rule based systems are presented in recent works. In [15] and [16], the rules are grouped into sets according to some criteria, and each rule within a set is statically compared to every other one to check consistency and completeness properties. In [17], within an exhaustive computationally expensive approach any chaining of rules is taken into account from which an inconsistency could be deduced. Despite [17], there are incremental approaches that check the rule base after each modification during development [18]. Some references use some graphical notations such as Petri nets to represent rules and detect the structural errors of rule bases. In [19], based on the concept of ω -nets [20], a special reachability graph is presented to detect structural errors in rule-based systems. This technique is applied in later researches [1, 14]. In [21], a fuzzy rule base systems verification method using high-level Petri nets is discussed. Furthermore, in [1], a double-phase verification technique for HBM is presented that consists of weak and strong verification whereby [14] adds a solution to semantically validate the HBM. In summary, the existing V&V techniques for rule bases mainly focus on structural verification and rarely deal with validation issues, let alone semantic validation of special cases such as human behavior models. Furthermore, in [22, 23], general V&V problems of human behavior models and its possible techniques are illustrated.

As a rather new field in human behavior, the educational system is regarded as the case-study, with which we deal in everyday life. In pursue to some recent works on student's performance evaluation [24, 25], the idea of entering the Professor into the system was initiated. In order to set up a rule base for the fuzzy system of human behavior, the effective parameters on decision making in that special field must be initially identified. On the way to this goal during careful consults with anthropology experts for a long time, many procedures were suggested to identify these parameters, among which lies survey research, -a subset of which is the Delfi technique- action research and correlation research. Finally according to the experts' recommendation, based on the techniques of questionnaire designation [26, 27 and 28], a

questionnaire was designed in a special way to be transformed to a fuzzy rule-base with uncertain fuzzy rules dealing with certainty factors. Then the specially designed questionnaire was handed among many students several times and after each time the necessary changes was made on it in order to mostly satisfy the students' (SME's) points of view. Then the constructed fuzzy rule base is mapped to a fuzzy Petri net and afterwards the corresponding special reachability graph is generated and searched in order to distinguish errors dealing with verification. Then by means of a rule referent gathered from the subject matter experts' (SME) point of view-here the student's point of view- the rule base is semantically validated.

The paper is organized as follows: section II deals with the errors concerning human behavior models. Section III concerns itself with the introduction of fuzzy Petri nets and mapping the rule base to fuzzy Petri net. Section IV is dedicated to the introduction of the case study. Section V and VI respectively illustrate the V&V processes. Section VII concludes the paper and presents ideas for future works.

II. FUZZY PETRI NETS

A. Fuzzy Petri Nets- a brief Introduction

A fuzzy Petri net model (FPN) can be used to represent a fuzzy rule-based system. A FPN [19, 29] is a directed graph containing two types of nodes: places and transitions, where circles represent places and bars represent transitions. Each place represents an antecedent or consequent and may or may not contain a token associated with a truth degree between zero and one which speaks for the amount of trust in the validity of the antecedent or consequent. Each transition representing a rule is associated with a certainty factor value between zero and one. The certainty factor represents the strength of the belief in the rule. The relationships from places to transitions and vice versa are represented by directed arcs. The concept of FPN is derived from Petri nets. As with [14 and its refs], a generalized FPN structure can be defined as an 8-tuple:

FPN = (P, T, D, I, O, μ , α , β), where

P = {p₁, p₂, ..., p_n} is a finite set of places,

T = {t₁, t₂, ..., t_m} is a finite set of transitions,

D = {d₁, d₂, ..., d_n} is a finite set of propositions ,

$P \cap T \cap D = \emptyset$, |P| = |D|

I: P × T → {0, 1} is the input function, a mapping from places to transitions;

O: T × P → {0, 1} is the output function, a mapping from transitions to places,

μ : T → [0, 1] is an association function, a mapping from transitions to [0, 1] i.e. the certainty factor

α : P → [0, 1] is an association function, a mapping from places to [0, 1] i.e. the truth degree

β : P → D, is an association function, a mapping from places to propositions.

B. Mapping The Rule Base to FPN

During this mapping procedure, each rule is represented as a transition with its corresponding certainty factor and each antecedent is modeled by an input place and the consequents are modeled by out places with corresponding truth degrees. In this modeling a transition- here a rule- is enabled to be fired if all its input places have a truth degree equal to or more than a predefined threshold value [30]. As illustrated in Fig.1, after firing the rule, the output places will have a truth degree equal to the input place truth degree multiplied by the transition certainty factor [30].

In order to transform compound rules to FPNs, we first apply normalization rules introduced in [21] to change any rules to Horn Clauses [21 and its refs]. A Horn Clause is a kind of rule in the following form.

$$P_1 \wedge P_2 \wedge P_3 \wedge \dots P_{j-1} \rightarrow P_j$$

III. ERRORS CONCERNING HUMAN BEHAVIOR MODELS

Human behavior models may suffer from two types of errors [14]. If modeled as a rule-base, they may suffer from the structural errors from which any fuzzy rule-base may suffer, among which incompleteness, inconsistency, circularity and redundancy are the most popular. On the other hand, HBMs as models being in contact with human operators and users, must meet the user's need. Any contrast with the user's point of view will be indicating the semantic errors dividing into two groups: semantic incompleteness and semantic incorrectness.

A. Structural errors

As illustrated in [14, 21] the structural errors from which a rule base may suffer are as follows.

1) Incompleteness

Incompleteness rules result from missing rules in a rule base. An example of incompleteness rules is as follows.

$$\begin{aligned} r_1: & \rightarrow P_1 \\ r_2: & P_1 \wedge P_3 \rightarrow P_2 \\ r_3: & P_1 \rightarrow P_4 \\ r_4: & P_2 \rightarrow . \end{aligned}$$

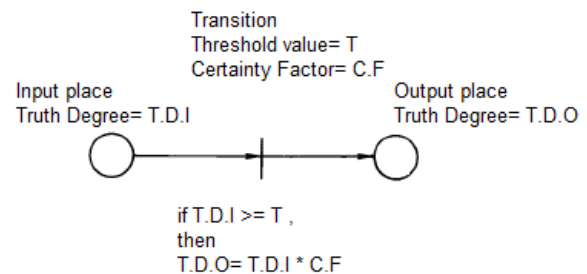


Figure 1. Firing process

Rule r₁ is representing fact -a source transition in the FPN, while rule r₄ is representing a query- a sink transition in the FPN. Rule r₂ is a useless rule because the antecedent P₃ does not have a matching part appearing in the consequents of the rest of the rules, thus P₃ is a dangling antecedent. Rule r₃ is a useless rule because the consequent P₄ of r₃ does not have a

matching part appeared in the antecedents of the rest of the rules, thus p_4 is a dead-end consequent.

2) Inconsistency

Inconsistency rules end in conflict and should be removed from the rule base. This means a set of rules are conflicting if contradictory conclusions can be derived under a certain condition. An example of inconsistency rules is as follows.

$$\begin{aligned} r_1: P_1 \wedge P_2 &\rightarrow P_3 \\ r_2: P_3 \wedge P_4 &\rightarrow P_5 \\ r_3: P_1 \wedge P_2 \wedge P_4 &\rightarrow \sim P_5 \end{aligned}$$

Rule r_3 is an inconsistent one because P_1 and P_2 ends in P_3 , P_3 and P_4 ends in P_5 , while in r_3 , P_1 and P_2 (the same P_3) and P_4 ends in $\sim P_5$.

3) Circularity

Circular rules refer to the case that several rules have circular dependency. Circularity may end in an infinite reasoning loop and must be broken. An example of circular rules is as follows.

$$\begin{aligned} r_1: P_1 &\rightarrow P_2 \\ r_2: P_2 &\rightarrow P_3 \\ r_3: P_3 &\rightarrow P_1 \end{aligned}$$

4) Redundancy

Redundancy rules are unnecessary rules in a rule base. Redundancy rules increase the size of the rule based and may cause extra useless deductions. An example of redundancy rules is as follows.

$$\begin{aligned} r_1: P_1 \wedge P_3 &\rightarrow P_2 \\ r_2: P_1 \wedge P_3 &\rightarrow P_2 \\ r_3: P_1 &\rightarrow P_2 \\ r_4: P_4 &\rightarrow P_5 \\ r_5: P_4 &\rightarrow P_5 \wedge P_6 \end{aligned}$$

r_1 is the redundant rule of r_2 . There are two cases of the directly subsumed rules. First, rules r_1 or r_2 is a subsumed rule of r_3 because r_1 or r_2 has more restrictive condition than r_3 . Second, rule r_4 is a subsumed rule of r_5 because r_4 has less implied conclusion than r_5 .

B. Semantic errors

As explained in [14] the semantic errors are classified into two levels.

1) Semantic incompleteness

In a human behavior model, semantic incompleteness happens if human behavior model does not meet users' requirements, and are reflected as missing rules, and missing antecedents or consequents in a rule from the users' point of view.

2) Semantic incorrectness

Semantic incorrectness occurs if the human decision model produces an output that is different from the expected output for given identical input data in the validation referent. Semantic incorrectness also indicates that the human behavior model doesn't meet the users' needs.

IV. MODELING OF THE CASE- STUDY

In recent works [1, 14], V&V process is applied on a ready given rule-base. In this work we are first supposed to construct a rule-base and then apply the V&V process on it.

In order to make up a rule base for a human related case-study. The needed information is gathered through some filled in questionnaires whose questions specially designed in two stages that leads to a hierarchical fuzzy inference within two steps. of course it seems necessary to notify that the students who were supposed to fill in the questionnaires were not informed of the mentioned structure lying beyond.

The questions that form the first step of the hierarchical deduction are in fact the input properties of the fuzzy HBR system whose outcome serves the internal properties of the system. Finally in the second stage, the fuzzy deduction on the internal properties and some directly effective input properties makes up the second stage rule.

The questions standing for the input variables supposed to make a unique internal property are arranged in a diversified form in order not to impose any conditional effect on the participant's mind while answering the questions and to let the students answer the questions feeling absolutely free and keeping the system as fuzzy as possible.

The answers to the questions is given within a 5-optional list which includes selecting a linguistic variable (very low, low, medium, high, very high) of the fuzzy system. Also to get the truth degree of the antecedents and consequents and also the certainty factor for each rule, the participants are asked to answer the following question in terms of percent: *How much you are confident to your answer?*

Our present case study is defined on the decision factors a student considers while selecting a professor among others for a special presented course by several professors.

A. The Input Properties

The system's input properties are gathered within 11 questions in the questionnaire. Questions 1 to 11 will be represented as Q1 to Q11 in the rest of paper. These questions are summarized as follows.

- Question 1-The professor's authority on the topic and his power to answer the questions?
- Question 2-Using references updated each semester and introducing further references to study?
- Question 3-How much the professor acts strong on presenting the topic?
- Question 4-How much he feels responsible for attending on time?
- Question 5-How much he optimally manages the time he has?
- Question 6- How much he appreciates the suggestions and constructive critics?
- Question 7-How much well-behaved he is?

- Question 8- Feeding the student with several quizzes and midterm exams?
- Question 9- presenting projects and assigning homework?
- Question 10- The range of marks?
- Question 11- Advice and suggestion by the elder students?

B. The Internal Properties

The internal properties of the system are made on the basis of some composition of the input properties.

- The input properties Q1, Q2 and Q3 form an internal property called “The power of teaching”.
- The input properties Q4 and Q5 form an internal property called “regularity”.
- The input properties Q6 and Q7 form an internal property called “behavior”.
- The input properties Q8 and Q9 form an internal property called “The power of attracting the student”.
- The power of teaching, regularity, behavior, the power of attracting the student, elders’ advice and the mark range are the factors making up the Prof’s rank to be selected among others.

In other words, we have a fuzzy deduction in two levels: Level one is supposed to deduce the internal properties, level two is supposed to deduce the prof’s rank based on the internal and input properties.

1) Level 1:

If Q1 is ... and Q2 is ... and Q3 is ..., then the power of teaching is ...

If Q4 is ... and Q5 is ..., then regularity is ...

If Q6 is ... and Q7 is ..., then behavior is

If Q8 is ... and Q9 is ... , then the power of attracting the student is

2) Level 2:

If the power of teaching is ...and regularity is ... and behavior is and attractiveness is ... then the prof’s popularity is

Each of the blanks is filled with a linguistic value: very low, low, medium, high and very high. A sample rule base for the above case-study can be presented as a human behavior model (HBM) as the following structure shown in Fig. 2.

- HBM=(Prof-Student, IPS, InPS, OPS, RS);
- HBM.IPS={Q1, Q2, Q3, Q4, Q5, Q6, Q7, Q8, Q9, Q10, Q11};
- HBM.InPS={Tea, Reg, Beh, Att};
- HBM.OPS={Pop};
- HBM. RS={R1, R2, ..., R10}

- HBM.RS.R1=(Rule1, Q1(vh) \wedge Q2(h) \wedge Q3(vh), Tea(vh), 0.95);
- HBM.RS.R2=(Rule2, Q4(h) \wedge Q5(h), Reg(h), 0.65);
- HBM.RS.R3=(Rule3, Q6(m) \wedge Q7(h), Beh(h), 0.85);
- HBM.RS.R4=(Rule4, Q8(m) \wedge Q9(h), Att(m), 0.65);
- HBM.RS.R5=(Rule5, Tea(vh) \wedge Reg(h) \wedge Beh(h) \wedge Att(m) \wedge Mark(m) \wedge Adv(h), Pop(h), 0.95);
- HBM.RS.R6=(Rule6, Q1(m) \wedge Q2(vh) \wedge Q3(h), Tea(m), 0.80);
- HBM.RS.R7=(Rule7, Q4(m) \wedge Q5(h), Reg(vh), 0.6);
- HBM.RS.R8=(Rule8, Q6(l) \wedge Q7(h), Beh(vh), 0.85);
- HBM.RS.R9=(Rule9, Q8(m) \wedge Q9(h), Att(l), 0.75);
- HBM.RS.R10=(Rule10, Tea(m) \wedge Reg(vh) \wedge Beh(vh) \wedge Att(l) \wedge Mark(vh) \wedge Adv(m), Pop(vh), 0.7);

In the above structure, human behavior model (HBM) is introduced within a 5-tuple consisting of the input property set (IPS), internal property set (InPS), output property set (OPS) and rule set (RS). Q1 to Q11 speak for Question 1 to Question 11 as input properties. Tea, Reg, Beh , Att and Pop respectively stand for the power of teaching, regularity, behavior, attractiveness and popularity as internal properties. Terms *vl*, *l*, *m*, *h* and *vh* respectively represent the linguistic values: very low, low, medium, high and very high. In the rules, the 2nd element shows the antecedents, the 3rd element shows the consequent and the last number shows the certainty factor dedicated to the rule. For example Rule1 is as follows.

- --- HBM.RS.R1=(Rule1, Q1(vh) \wedge Q2(h) \wedge Q3(vh), Tea(vh), 0.95);
- If Q1 is very high and Q2 is high and Q3 is very high, then the power of teaching is very high.

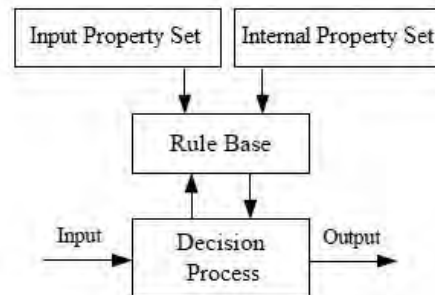


Figure 2. The decision model

The corresponding Petri net model is illustrated in Fig. 3. In this Petri net model, according to the propositions dedicated to each place, transitions 1 to 10 respectively represent rules 1 to 10 in the introduced rule base above and firing each transition means the corresponding rule is fulfilled.

V. VERIFICATION PROCESS

In order to fulfill the rule-base verification phase, we must first map the rule-base to Petri net as shown in Fig. 3. Then as with the algorithm mentioned in [19, 21] a special reachability graph is generated on the basis of the concept of ω -nets.

In this reachability graph, first, a zero vector is defined as the root node as long as the number of places. Then at any current marking, among the transitions yet not considered, the enabled transitions are determined. At each step by firing the set of enabled transitions, a new node is added to the graph in which the corresponding elements of the node- the places which are filled after firing the transitions- are set to ω which is assumed as a huge value. In this way at each step there's a marking. If firing of the transitions at a step ends in a repetitive marking, the graph will have a loop.

After generating the reachability graph, the structural errors including incompleteness, inconsistency, redundancy and circularity are distinguished. Then on the basis of the

SME point of view, the rule-base is verified to eliminate the errors.

The corresponding reachability graph for the above Petri net model is depicted in Fig. 4. The places P0 to P17 are regarded as TRUE antecedents and are initially filled (set to ω) for this reason. That's why in the first node there are 18 ω 's. In this marking transitions T1, T2, T3, T4, T6, T7, T8 and T9 are enabled. After firing these transitions, in the second step, places P18 to P25 are filled and the corresponding values in the node vector are set to ω . On the final step by firing T5 and T10 (the enabled transitions), the places P26 and P27 will also be filled up.

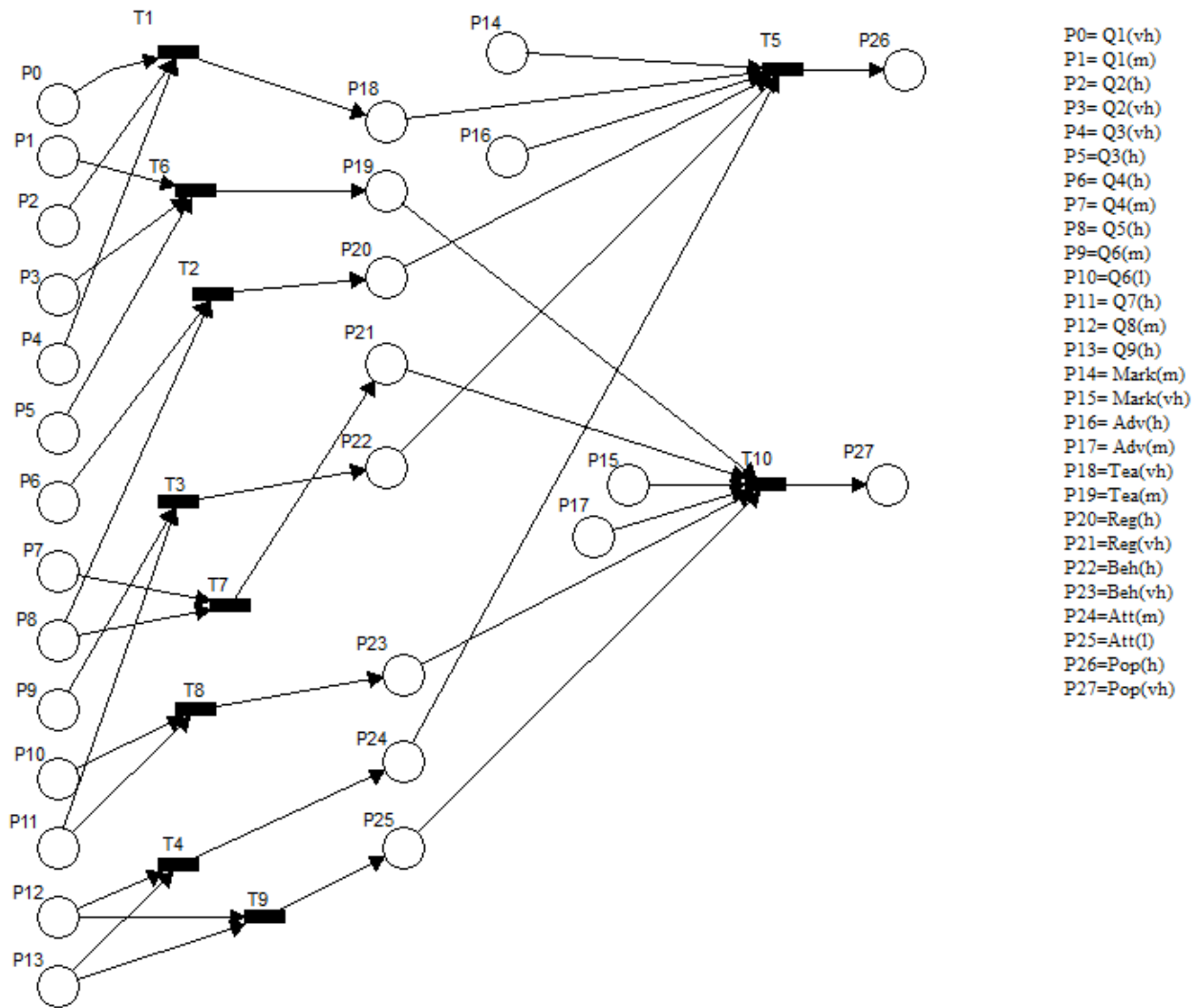


Figure 3. The Petri net representation of the HBM system

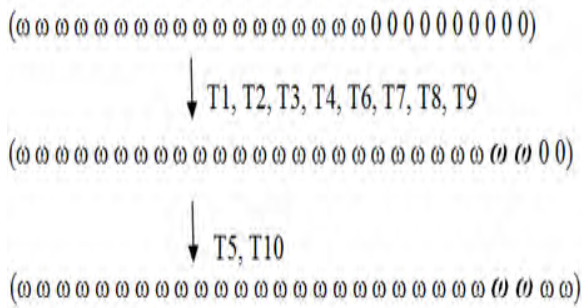


Figure 4. The reachability diagram

In the reachability graph shown in Fig. 4, all the places and transitions exist so there are no incompleteness errors. P24 and P25 are different states of one property (attractiveness). Their simultaneous existence may refer to some concept of inconsistency in the rule base. The reachability graph lacking any loops speaks for lacking circularity errors and finally having no transitions underlined, speaks for non-redundancy.

In order to verify the rule base after verification, the SME point of view must be considered to decide whether to omit or change a rule, or which rule must be changed or omitted to leave the rule base as a structurally fault-free one.

After a careful consult with the SME- here the students, we conclude that regarding the SME opinion, by omitting the Rule 9, the rule-base is refined after verification.

VI. VALIDATION PROCESS

In order to fulfill the validation process, first of all a validation referent is required. In this research the information needed to construct a validation referent is gathered from the students' point of view - the SME of the present system.

Validation is carried out within two separate phases: static validation and dynamic validation [14].

A. Static Validation

Static validation involves distinguishing the semantic incompleteness. In the static validation running or reasoning the FPN is not needed. In this phase only the places are searched and their properties are recorded and compared with the referent. If the number of searched input properties is less than the expected ones, the rule base may miss antecedents, if the number of searched output properties is less than the expected ones, the rule base may miss consequents and if the number of searched internal properties is less than the expected ones, the rule base may miss antecedents or consequents.

The referent for validation given by a student is as follows and the static validation results are summarized in Table1.

- HBM= (Prof-Student, IPSref, InPSref, OPSref, RSref);
- HBM.IPSref={Q1, Q2, Q3, Q4, Q5, Q6, Q7, Q8, Q9, mark, advice};
- HBM.InPSref={Tea, Reg, Beh, Att, scientific reputation};

- HBM.OPSref={Pop};
- HBM. RSref={Rref1, Rref2, ..., Rref10}
- HBM.RS.R1ref =(Rule1ref, Q1(vh) \wedge Q2(h) \wedge Q3(vh), Tea(vh), 0.95);
- HBM.RS.R2ref =(Rule2ref, Q4(h) \wedge Q5(h), Reg(h), 0.65);
- HBM.RS.R3ref =(Rule3ref, Q6(m) \wedge Q7(h), Beh(vh), 0.7);
- HBM.RS.R4ref =(Rule4ref, Q8(m) \wedge Q9(h), Att(m), 0.65);
- HBM.RS.R5ref =(Rule5ref, Tea(vh) \wedge Reg(h) \wedge Beh(h) \wedge Att(m) \wedge Mark(m) \wedge Adv(h), Pop(h), 0.95);
- HBM.RS.R6ref =(Rule6ref, Q1(m) \wedge Q2(vh) \wedge Q3(h), Tea(m), 0.80);
- HBM.RS.R7ref =(Rule7ref, Q4(m) \wedge Q5(h), Reg(h), 0.5);
- HBM.RS.R8ref =(Rule8ref, Q6(l) \wedge Q7(h), Beh(vh), 0.85);
- HBM.RS.R10ref =(Rule10ref, Tea(m) \wedge Reg(vh) \wedge Beh(vh) \wedge Att(l) \wedge Mark(vh) \wedge Adv(m), Pop(vh), 0.7);

According to Table1, the property of scientific reputation and rules R3ref and R7ref do not exist in the present rule-base. So the rule-base suffers from semantic incompleteness in this student's point of view. After interviewing many students, we conclude that the "scientific reputation" property can be neglected and the R7 is replaced with R7ref and R3 remains the same.

B. Dynamic Validation

Dynamic validation involves clarifying the existence of semantic incorrectness through running and reasoning FPN. In order to fulfill the dynamic validation, the results of reasoning FPN for given inputs are compared to their counterparts in the validation referent to check if there's any semantic incorrectness.

As with [30], rules with certainty factors are classified into three types, among which we use the first type according to the nature of the existing rules.

- Type 1: if P_1 and P_2 and ... P_n , then P_m .
- Type 2: if P_n then P_1 and P_2 ... and P_n .
- Type 3: if P_1 or P_2 ... or P_n , then P_m .

If α_i is considered as the truth degree of antecedents or consequents and μ_i is the certainty factor dedicated to rule r_i , the rule and its uncertainty reasoning is as follows.

$$R_i: P_1(\alpha_1) \wedge P_2(\alpha_2) \wedge \dots P_{n-1}(\alpha_{n-1}) \rightarrow P_n(\alpha_n) \quad CF = \mu_i$$

$$\alpha_n = \min\{\alpha_1, \alpha_2, \dots, \alpha_{n-1}\} \times \mu_i$$

In this part, the revised rule-base after the static validation must be considered.

TABLE I. THE STATIC VALIDATION RESULTS

Present FPN				Referent FPN			
IPS	InPS	OPS	RS-verified	IPSref	InPSref	OPSref	RSref
Q1	Teach	Pop	R1	Q1	Teach	Pop	R1ref
Q2	Regulation		R2	Q2	Regulation		R2ref
Q3	Behavior		x	Q3	Behavior		R3ref
Q4	Attractiveness		R4	Q4	Attractiveness		R4ref
Q5	x		R5	Q5	Scientific Reputation		R5ref
Q6			R6	Q6			R6ref
Q7			x	Q7			R7ref
Q8			R8	Q8			R8ref
Q9			R10	Q9			R10ref
Mark				Mark			
Advice				Advice			

The reference values for dynamic validation given by the student are as follows. The following numbers for reference values are gathered throughout the questionnaires by adding a choice to be filled in, in percent form.

-- reference values:

$$\text{Ref - value 1: } \alpha(Q1(vh)) = 0.65 \wedge \alpha(Q2(h)) = 0.75 \wedge \alpha(Q3(vh)) = 0.9 \rightarrow \alpha(\text{Tea}(vh)) > 0.7$$

$$\text{Ref - value 2: } \alpha(Q4(h)) = 0.7 \wedge \alpha(Q5(h)) = 0.8 \rightarrow \alpha(\text{Reg}(h)) > 0.4$$

$$\begin{aligned} \text{Ref - value 3: } & \alpha(Q1(vh)) = 0.65 \wedge \alpha(Q2(h)) \\ & = 0.75 \wedge \alpha(Q3(vh)) = 0.9 \wedge \alpha(Q4(h)) \\ & = 0.7 \wedge \alpha(Q5(h)) = 0.8 \wedge \alpha(Q6(m)) \\ & = 0.6 \wedge \alpha(Q7(h)) = 0.45 \wedge \alpha(Q8(m)) \\ & = 0.75 \wedge \alpha(Q9(h)) \\ & = 0.9 \wedge \alpha(\text{mark}(m)) \\ & = 0.95 \wedge \alpha(\text{adv}(h)) = 0.9 \rightarrow \alpha(\text{Pop}(h)) \\ & > 0.3 \end{aligned}$$

Considering the certainty factors in the validation referent given above, by the reference values for reasoning we have:

Ref-value 1 is validated by the use of Rule1ref according to the correspondence between their antecedents and consequents. Minimum of the truth degrees of the antecedents given in the referent i.e. $\min(0.65, 0.75, 0.9)$, must be multiplied to the certainty factor given in the referent rule base (0.95) to obtain the truth degree of the consequent and compare it with the condition provided by the referent. As illustrated underneath, for this case, the validation criterion fails.

$$\min(0.65, 0.75, 0.9) \times 0.95 = 0.6175 < 0.7$$

\rightarrow semantic incorrectness, validation failed

Similarly Ref-value 2 is validated by Rule2ref. Minimum of the truth degrees of the antecedents given in the referent i.e.

$\min(0.7, 0.8)$, is multiplied to the certainty factor given in the referent rule base (0.65) to obtain the truth degree of the consequent and is compared to the condition provided by the referent. For this case, the validation criterion is passed as follows.

$$\min(0.7, 0.8) \times 0.65 = 0.455 > 0.4$$

$\rightarrow o.k!$ passed validation.

In some cases the referent values maybe given in such a way that it is needed to merge rules during validation. In order to validate the Ref-value 3, according to the truth degrees given for special antecedents, rules Rule1ref to Rule5ref must be merged during validation first to obtain the truth degrees for the antecedents of Rule5ref.

- validation through Rule1ref to obtain the truth degree for Tea(vh) gives: $\min(0.65, 0.75, 0.9) \times 0.95 = 0.6175$
- validation through Rule2ref to obtain the truth degree for Reg(h) gives: $\min(0.7, 0.8) \times 0.65 = 0.455$
- validation through Rule3ref to obtain the truth degree for Beh(vh) gives: $\min(0.6, 0.45) \times 0.7 = 0.315$
- validation through Rule4ref to obtain the truth degree for Att(m) gives: $\min(0.75, 0.9) \times 0.65 = 0.4875$
- and finally validation through Rule5ref to obtain the truth degree for Pop(h) gives:
 $\min(0.6175, 0.455, 0.315, 0.4875, 0.95, 0.9) \times 0.95 = 0.29925 < 0.3$
 \rightarrow semantic incorrectness, validation failed

From the above, it can be concluded that the rule base suffers from semantic incorrectness in the point of view of the student who provided the reference values. However if the difference between 0.61 and 0.7 and also between 0.299 and 0.3 is neglectable, we can conclude that the rule base is near the validation criteria.

VII. CONCLUSION AND FUTURE WORK

Improving the fidelity and automation of simulations, human behavior modeling is the concern of nowadays research. On this way, in this research a new case study dealing with professor student interaction is defined. The corresponding rule base was constructed by gathering information through specially designed questionnaires. The rule base was mapped to a FPN and through an FPN-based recently presented method was verified to distinguish and refine the structural errors. Afterwards, the semantic errors were distinguished by reasoning the FPN through the dynamic validation.

In the future the presented case-study system will be improved by inserting the professor and student's personal and cultural characteristics within beta-distributions. Also improving the optimality of the certainty factors and truth degrees by artificial intelligence algorithms is not far to imagine if these factors are defined as a fitness function of the student's characteristics such as age, sex, desire for PhD, ranking among classmates. Furthermore the threshold values for transition enabling can be adjusted on the basis of the results of the fuzzy reasoning. Also the concept of truth degrees which dedicate a value to the tokens inside the places may initiate the idea of using colored Petri nets with valued tokens. Also considering priority to fire the enabled transitions in the reachability graph may make us set foot on priority Petri nets.

REFERENCES

- [1] Fei Liu, Ming Yang, Guobing Sun, "Verification of Human Decision Models in Military Simulations", Proceedings of the IEEE 2007, The First Asia International Conference on Modeling & Simulation, pp. 363 - 368
- [2] B. H. McNally, "An approach to human behavior modeling in an air force simulation", in Proc. IEEE 2005, Winter Simulation Conference, Orlando, Dec. 2005, pp. 1118-1122.
- [3] Naoyuki Kubota, and Kenichiro Nishida, "Prediction of Human Behavior Patterns based on Spiking Neurons", The 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN06), Hatfield, UK, September 6-8, 2006.
- [4] Zoran Duric, Wayne D. Gray, Ric Heishman, Fayin Li, Azriel Rosenfeld, Michael J. Schoelles, Christian Schunn, and Harry Wechsler, "Integrating Perceptual and Cognitive Modeling for Adaptive and Intelligent Human-Computer Interaction", Proceedings of the IEEE, Vol. 90, No. 7, pp. 1272 - 1289, July 2002.
- [5] Hiroyuki Okuda, Soichiro Hayakawa, Tatsuya Suzuki and Niuo Tsuchida, "Modeling of Human Behavior in Man-Machine Cooperative System Based on Hybrid System Framework", 2007 IEEE International Conference on Robotics and Automation Roma, Italy, 10-14 April 2007.
- [6] Wei Ding, Lili Pei, Hongyi Li, Ning Xi, Yuechao Wang, "The Effects of Time Delay of Internet on Characteristics of Human Behaviors", Proceedings of the 2009 IEEE International Conference on Networking, Sensing and Control, Okayama, Japan, March 26-29, 2009.
- [7] David W. Dorsey and Michael D. Coovert, "Mathematical Modeling of Decision Making: A Soft and Fuzzy Approach to Capturing Hard Decisions", HUMAN FACTORS, Vol. 45, No. 1, Spring 2003, pp. 117-135.
- [8] Michael J. Johnson, Michael McGinnis, "methodology for human decision making using using fuzzy artmap neural networks", Proceedings of the 2002 International Joint Conference on Neural Networks, Vol.3, pp. 2668 - 2673, 2002
- [9] Simon C.K. Shiu, James N.K. Liu, Daniel S. Yeung, "An Approach Towards the verification of Fuzzy Hybrid Rule/Frame-based Expert Systems", 12th European Conference on Artificial Intelligence, Published in 1996 by John Wiley & Sons, Ltd
- [10] Zhang Hua, Li Rui, Sun Jizhou, "An Emotional Model for Nonverbal Communication based on Fuzzy Dynamic Bayesian Network", Canadian Conference on Electrical and Computer Engineering, pp. 1534 - 1537 2006.
- [11] Juanda Lokman Jun-ichi Imai Masahide Kaneko, "Understanding Human Action in Daily Life Scene based on Action Decomposition using Dictionary Terms and Bayesian Network", IEEE 2008 Second International Symposium on Universal Communication.
- [12] Wang Li and Wang Mingzhe, "Extraction and Confirmation of Rules for Human Decision Making", IEEE 2009 International Forum on Information Technology and Applications.
- [13] Zhuomin Sun, "Multi-Agent Based Modeling: Methods and Techniques for Investigating Human Behaviors", Proceedings of the 2007 IEEE International Conference on Mechatronics and Automation August 5 - 8, 2007, Harbin, China.
- [14] Fei Liu, Ming Yang and Peng Shi, "Verification and Validation of Fuzzy Rules-Based Human Behavior Models", 7th International Conference on System Simulation and Scientific Computing, 2008, pp. 813 - 819
- [15] M. Suwa, A. Scott, and E. Shortliffe, "An approach to verifying completeness and consistency in a Rule-Based expert system," Technical Report: CS-TR-82-922, pp. 16-21, 1982, Stanford University, Stanford, CA, USA
- [16] T. Nguyen, W. Perkins, Y. Laffey, and D. Pecora, "Checking an expert systems knowledge base for consistency and completeness," in Proc. International Joint Conference on Artificial Intelligence, 1985, pp. 375-378.
- [17] M. Toussot, "On the consistency of knowledge bases: the COVADIS system", in Proc. European Conference on Artificial Intelligence, 1988, pp. 79-84.
- [18] P. Meseguer, "Incremental verification of rule-based expert systems," in Proc. European Conference on Artificial Intelligence, 1992, pp. 840-844.
- [19] He, X., Chu, W. C., Yang, H., Yang, S. J.H. "A New Approach to Verify Rule-Based Systems Using Petri Nets". In: IEEE proceeding of 23th Annual International Computer Software and Applications Conference (COMPSAC'99). (1999) 462-467.
- [20] T. Murata, "Petri Nets: Properties, Analysis and Application," Proc. IEEE, vol. 77, no. 4, pp. 541-580, 1989.
- [21] S. J. H. Yang, J. J. P. Tsai, and C. Chen, "Fuzzy rule base systems verification using high-level Petri nets," IEEE Transactions on Knowledge and Data Engineering, vol. 15, no. 2, pp. 457-473, Mar./Apr. 2003.
- [22] A. J. Gonzalez, and M. Murillo, "Validation of human behavioral models," in Proc. Simulation Interoperability Workshop, Mar. 1999.
- [23] S. Y. Harmon, and S. M. Youngblood, "Validation of human behavior representations," in Proc. Simulation Interoperability Workshop, Mar. 1999.
- [24] Hui-Yu Wang and Shyi-Ming Chen, "Evaluating Students' Answer scripts using Fuzzy Numbers Associated With Degrees of Confidence", IEEE Transactions on Fuzzy Systems, Vol. 16, No. 2, April 2008
- [25] Sunghyun Weon And Jinil Kim, "Learning Achievement Evaluation Strategy using Fuzzy Membership Function", 31st ASEE/IEEE Frontiers in Education Conference, October 10 - 13, 2001.
- [26] GU Dong-xiao, LIANG Chang-yong, CHEN Wen-en, GU Ya-di, FAN Xin, WU Wei, "Case-based Knowledge Reuse Technology for Questionnaires Design", 4th International Conference on Wireless Communications, Networking and Mobile Computing, 2008, pp. 1 - 4
- [27] I. R. Craig and G. L. Burrett, "The Design Of A Human Factors Questionnaire For Cockpit Assessment", An International Conference on

- Human Interfaces in Control Rooms, Cockpits and Command Centres, 21 - 23 June 1999
- [28] Ayushi Garg and Sumit Singh, "Towards The Adaptive Questionnaire Generation using Soft Computing", World Congress on Nature & Biologically Inspired Computing, 2009, pp. 806 - 811
- [29] Shen, V. R. L.: Knowledge Representation using High-Level Fuzzy Petri Nets. IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems And Humans. Vol. 36, No. 6. (2006) 1220-1227
- [30] S.M. Chen, J.S. Ke, and J.F. Chang, "Knowledge Representation using Fuzzy Petri Nets," IEEE Trans. Knowledge and Data Eng., vol. 2, no. 3, pp. 311-319, Sept. 1990.

SVD-EBP Algorithm for Iris Pattern Recognition

Mr. Babasaheb G. Patil

Department of Electronics Engineering
Walchand College of Engineering,
Sangli, (Maharashtra) India

Dr. Mrs. Shaila Subbaraman

Department of Electronics Engineering
Walchand College of Engineering,
Sangli, (Maharashtra) India

Abstract— This paper proposes a neural network approach based on Error Back Propagation (EBP) for classification of different eye images. To reduce the complexity of layered neural network the dimensions of input vectors are optimized using Singular Value Decomposition (SVD). The main objective of this work is to prove usefulness of SVD to form a compact set of features for classification by EBP algorithm. The results of our work indicate that optimum classification values are obtained with SVD dimensions of 20 and maximum number of classes as 9 with the state-of-the art computational resources. The details of this combined system named as SVD-EBP system for Iris pattern recognition and the results thereof are presented in this paper.

Keywords- Singular value decomposition (SVD); Error back Propagation (EBP).

I. INTRODUCTION

Biometrics is one of the areas of research that has gained widespread acceptance in the field of human identification and fraud prevention. Although the current state-of-the-art provides reliable automatic recognition of biometric features, the field is not completely researched. Different biometric features offer different degrees of reliability and performance. The Human iris is one of the biometric parameters of the human body efficiently used for person recognition. Number of researchers has worked on person identification using iris as biometric [1, 2, and 3]. However the results obtained by them indicate strong dependence of recognition/classification accuracy on the orientation of the iris image as well as on the light intensity levels while capturing images.

This paper confirms the usefulness of Singular Value Decomposition method- SVD (as was recommended by [4]) to extract a compact set of features from the preprocessed iris image to overcome the earlier drawback of orientation and intensity dependent classification accuracy for iris recognition.

A. Characteristics of Human Iris [4]

The use of the human iris as a biometric feature offers many advantages over other human biometric features. The Iris is the only internal human body organ that is visible from the outside and is well protected from external modifiers. A fingerprint, for example, may suffer transformations due to harm or aging, voice patterns may be altered due to vocal diseases. However, the human iris image is relatively simple to acquire and may be done so in a non-intrusive way. The Human iris starts forming right from the third month of gestation in the mother's uterus. A small part of final iris pattern is developed from the individual DNA while most of

the part is developed randomly by the growth of epithelial tissues present there. It means that two eyes from the same individual, although they look very similar, have two different patterns of two Irises, however with unique DNA related internal pattern. Identical twins would then exhibit four different iris patterns and can be uniquely recognized using these patterns. Hence human identification using iris images has become a favorite choice among researchers in recent years.

II. GENERIC PROCESS FLOW

A generic process flow for iris pattern recognition and classification is as shown in Figure 1. As it is clear from the figure, the entire process flow broadly consists of two parts. The first part corresponds to image pre-processing in order to extract an optimal and compact set of features while the second part deals with pattern recognition and classification. In our research work a Singular Value Decomposition (SVD) method was used to extract a compact set of features on the pre-processed iris images where the pre-processing was carried out using standard steps viz. image acquisition, image segmentation, edge detection etc.

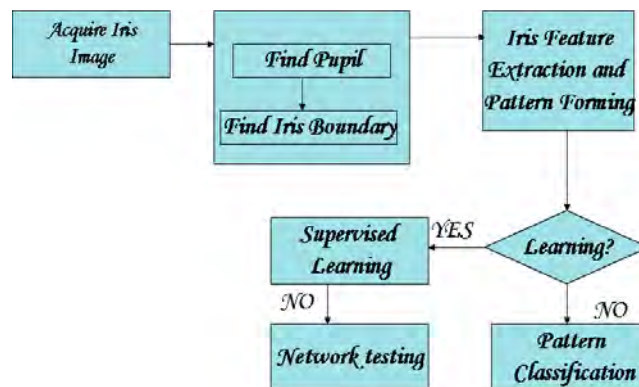


Figure.1: Processing Steps

In the second part, these features were inputted to multilayered neural network classifier implementing Error Back Propagation (EBP) algorithm. EBP is a supervisory algorithm generally used in two phases' viz. training phase and testing phase. MATLAB was extensively used in the entire research work. The details of preprocessing are given in the next section while the details of recognition and classification are presented in the Section IV of the paper. The conclusions derived from this work are presented in the Section V highlighting the future scope in this area.

III. PREPROCESSING

As stated above, pre-processing contains following three steps [6, 7]

- Image Acquisition
- Image Segmentation and Edge Detection
- Feature Extraction

The details of these steps are given below.

A. Image Acquisition

The major problem of iris Recognition using already researched methods is image acquisition because of the susceptibility of eyes to degree of illumination. The importance of the image grabbing system implementing consistent illumination has been spelled out in literature. But the approaches of SVD-EBP used by the authors of this paper do not pose any such restriction on the intensity level of light illumination. However, the pupil is an open door to the retina, one of the most sensitive organs of our body, and extra care must be taken when shedding direct light on it.

The work presented in this paper uses the CASIA iris database as input.[8] This database uses a special camera that operates in the infrared spectrum of light, not visible by the human eye. Here, each iris class is composed of 7 samples taken in two sessions, three in the first session and four in the second session. The two sessions were taken with an interval of one month. Images are 320x280 pixels gray scale taken by a digital optical sensor designed by NLPR (National Laboratory of Pattern Recognition Chinese Academy of Sciences). There are 108 classes or total number of iris images is 756. Figure 2 shows a sample of the image of an eye from this database.

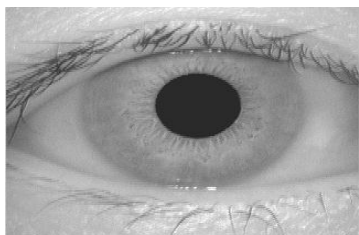


Figure.2. Image of an Eye

B. Iris Segmentation

The main motive behind iris Segmentation is to remove the non-useful information like the sclera and the pupil information and extract the region of interest. First, the pupil is detected and then the iris-sclera boundary is detected which is usually done in two steps B1 and B2 as given below [4].

1) Detection of Pupillary Boundary

As we all know that pupil is a very dark blob of a certain minimum size in the picture and no other segment of continuous dark pixels is of the same size. This makes an easy task to detect the pupillary boundary in the iris image. This algorithm finds the center of the pupil and two radial coefficients as the pupil is not always a perfect circle.

To find the pupil, we first need to apply a step threshold to the image given by,

$$g(x) = \begin{cases} f(x) > 70: 1 \\ f(x) \leq 70: 0 \end{cases}$$

where $f(x)$ is the original image and $g(x)$ is the threshold image. Pixels with intensity greater than the empirical value of 70 (in a 0 to 255 scale) are dark pixels, therefore converted to 1 (white). Pixels with intensity smaller than or equal to 70 are assigned to 0 (black). Figure 3 shows the threshold image of the pupil. Since, the eyelashes also satisfy the threshold condition, they are visible in the figure.



Figure.3. Threshold Image

To eliminate the area of eyelashes, search of a region of 8 connected pixels with value 1 is carried out. The report on CASIA database indicates that an area value of 2500 is sufficient for pupil region. The eyelashes definitely have much smaller region than the pupil region, hence the area associated with eyelashes would be much smaller than 2500. Using this knowledge, one can cycle through all regions and apply the following condition:

for each Region R
if $AREA(R) < 2500$
set all pixels of R to 0

Thus the pupil is separated and the centroid (x_{cp} , y_{cp}) of the pupil is extracted. Also horizontal and vertical radii are calculated. Figure 4 shows the threshold image in which the eyelashes have been cropped out by Freeman's Chain Coding method [4].

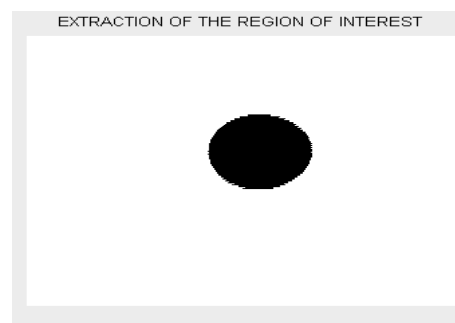


Figure.4. Image with the Eyelashes Cropped out.

2) Iris Edge Detection

Figure.5 shows the pupil in the original eye image with centroid (x_{cp} , y_{cp}), horizontal radius r_x and vertical radius r_y .

After detecting the pupil the next step is to find the contour of the iris. Already, we have detected the pupil location and we have the knowledge that it is concentric to the outer perimeter of the iris.

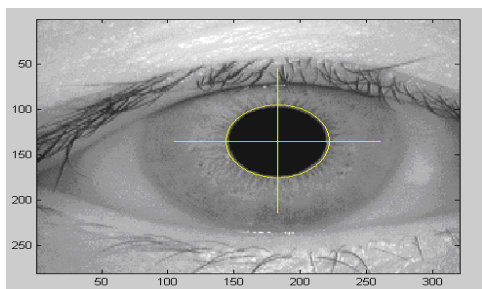


Figure.5 . Detection of the centroid (x_{cp}, y_{cp}) and horizontal radius r_x and vertical radius r_y

But, the problem is that sometimes the eyelid may occlude part of the iris. Also, the iris center may not match with the pupil center, and we will have to deal with strips of iris of different width around the pupil. This method takes into consideration the fact that areas of the iris at the right and left side of the pupil contain the most significant information that is useful for data extraction. The areas above and below the pupil carry unique information, but it is very common that they are totally or partially occluded by eyelashes or eyelids.

Figure 6 shows the steps how to find the right edge of the iris. The strategy adopted for iris detection is to trace a horizontal imaginary line that crosses the whole image passing through the center of the pupil [4]. This figure shows the horizontal line passing through y_{cp} (center of pupil) of original image. Corresponding to pixels on this line, the pixel intensities are also shown in the figure graphically.

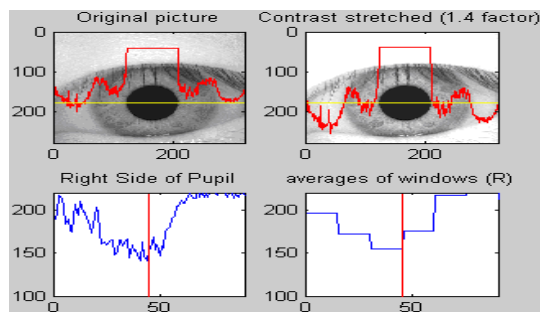


Figure 6: Iris Edge Detection

Starting from the edges of the pupil, we analyze the signal composed by pixel intensity from the center of the image towards the border and try to detect abrupt increases of intensity level. Although the edge between the iris disk and the sclera is most of the times smooth, it is known that it always has greater intensity than iris pixels. We intensify this difference applying a linear contrast filter. It is possible that some pixels inside the iris disk are very bright, causing a sudden rise in intensity. That could mislead the algorithm to detect it as an iris edge at that point. To avoid this, we take the average intensity of small windows when the sudden rises occur from these intervals as shown in Figure 6.

C. Feature Extraction

Once the segmentation has been performed and the region of interest (ROI) has been extracted, the next step is to extract the features so as to reduce the problem of dimensionality. The iris Basis Images are extracted by converting polar information of the iris into Cartesian information with radial resolution of 10 pixels and specific angular resolution. in terms of pixel

count (Y-axis) for angles varying from 0^0 to 360^0 (X-axis) in specific steps. Pixels on either side of the pupil are collected and one reduced image of the iris is formed. Thus, the extraction of iris basis reduces the dimension as the non-useful information is cropped out. We can better visualize this strategy by looking at Figure 7 as shown below.

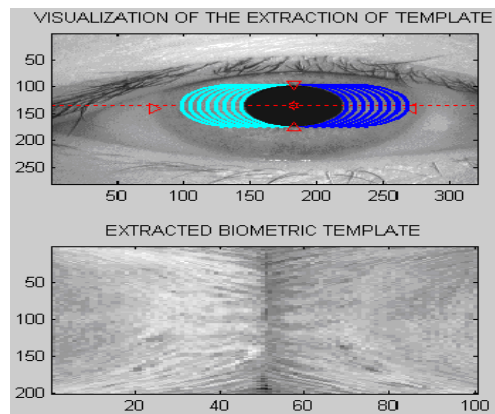


Figure.7 Extraction of the Biometric Template

Up-till now though the dimension has been reduced significantly, it is still too high for classification. So, the singular value decomposition, as explained below, is employed for further reduction in dimension.

1) Singular Value Decomposition (SVD)

Singular Value Decomposition is a powerful tool in matrix computations and analyses which has the advantage of being quite robust to numerical errors [9,10]. Additionally SVD exposes the geometric structure of full-rank matrix. The presence of noise in image, either from image capturing systems or from round-off numerical errors, results into an image matrix that is generally of full rank. Hence SVD is a recommended technique to decompose the data into an optimal estimate of signal and the noise components. Further SVD aids in image compression by storing the image information in only M elements of a $M \times N$ image matrix ($M \geq N$). For all these reasons, we have used SVD in deriving a compact and representative set of features of the iris images grabbed even with light intensity variation and in presence of noise.

The basic operation of SVD relies on the factorization of an $M \times N$ matrix ($M \geq N$) into three other matrices on the following form:

$$A = U^T \delta V$$

where the superscript “ T ” denotes transpose. U is an $M \times M$ orthogonal matrix, V is an $N \times N$ orthogonal matrix and δ is an $M \times N$ diagonal matrix with $s_{ij} = 0$ if $i \neq j$ and $s_{ii} \geq s_{i+1,i+1}$. The two important aspects to be noted here are:

1. δ is zero everywhere except in the main diagonal. This leads to reduction in the dimension of the input pattern from a matrix $M \times N$ to only a vector of N elements.
2. Only the first few elements contain substantial information, and the vector tail without significant loss of information can be cropped out.

Figure 8 shows a plot of SVD pattern vectors which very well depicts the second property mentioned above.

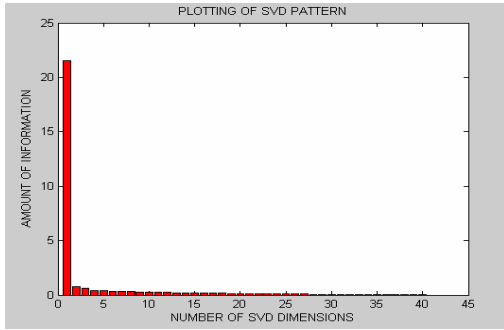


Figure.8: SVD Pattern Vectors.

IV. CLASSIFICATION

As mentioned in the Generic Process Flow, second part of this paper deals with iris classification. Here, the SVD features extracted as explained above were inputted to multilayered neural network classifier implementing Error Back Propagation (EBP) algorithm. Here, any five patterns of total 7 patterns of each of the 108 classes of CASIA database were used for network training and the remaining two patterns of each of 108 classes were used for network testing. Our network implements the classical 3-layer architecture: Input layer, Hidden layer and Output layer. The input layer contains as much neurons as the dimensionality of the pattern vector, which we have limited to 10 (Refer Figure 8) in the present case. The number of neurons in the hidden layer is approximately double as that of input layer (i.e. 20) for good classification results. The number of neurons in the output layer corresponds to the number of classes to be recognized.

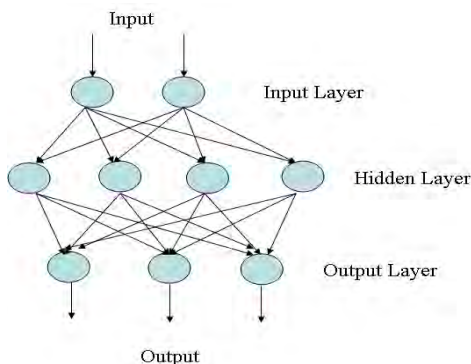


Figure.9. Architecture of Feed Forward Back propagation Neural Network

A. Network Design

Following are some parameters set for network training in MATLAB

- Training function: traingda (Adaptive learning rate)
- Initial learning rate: 0.2
- Learning rate increment: 1.05
- Maximum Epochs: 50,000
- Error goal: 5×10^{-7}
- Minimum gradient: 1×10^{-9}

The various experiments performed to train the neural network and test the iris pattern include iris-basis images of

40*40 pixels quantized from the original iris image with a mask of 3*3 pixels. The SVD algorithm as discussed in the section III of this paper was developed to output the vectors with 3, 10, 20 and 40 dimensions. The target classes for classification were varied from 3 to 20.

When the network was trained in the supervised mode, a target vector was also presented to the network. This target vector has every element set to zero, except on the position of the target class that will be set to 1. The idea behind this design decision is that for each input pattern X presented to the network, an output vector Y is produced. This vector has the number of elements equal to numbers of output neurons. Each output neuron implements a squashing function that produces a real number in the range (0, 1). To determine which class is being indicated by the network, we select the maximum number in Y and set it to 1, while setting all other elements to zero. The element set to one indicates the classification of that input pattern. Figure 10 shows the convergence behavior of neural network using Gradient Descent algorithm for a typical case as obtained from MATLAB. It is seen that the MSE was reached within 2150 epochs.

While simulating using MATLAB, it was found that as the number of classes increased, the network had more difficulty in learning the proper discriminatory weights. The network was able to reach the MSE goal within the specified number of epochs. For number of classes > 6, the MSE goal was not attained anymore, but the MSE kept decreasing until the maximum number of epochs was reached. We feel that increasing the number of epochs may allow the network to eventually converge.

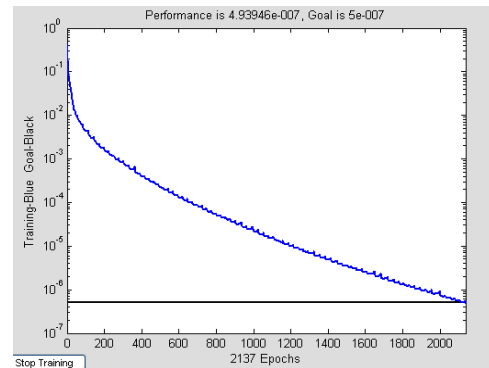


Figure10. Training of the Feed Forward Neural Network

V. RESULTS

Table I shows the classification accuracy with variation in the number of SVD dimensions and number of classes. It is seen from the table that the classification accuracy is 100% up-to five number of classes with requirement of at least 20 SVD dimensions. Increasing these dimensions to 40 does not make significant difference in the classification accuracy. But, it can be seen that as the number of classes are increased beyond 9 the system tends to become over-biased towards one class and the classification accuracy becomes poor even with forty SVD dimensions.

This behavior is also shown in Figure 11. Though the trend of our results using SVD-EBP system is consistent with that of

[4], the classification accuracies obtained by us for a specific class and specific dimension of SVD vector are better than those reported in this reference. From the results it is seen that this method of classification perhaps cannot be applied for classes above 20. This is due to the fact that with increase in number of classes, a huge computational burden is being exerted on neural network making it incapable to handle voluminous data in spite of the reduction in the dimensionality of the input vector (representative of 2-D iris image) obtained by SVD approach.

TABLE I. CLASSIFICATION OF DIFFERENT CLASSES USING DIFFERENT DIMENSION

Classification Accuracy				
Number of Classes	Number of SVD Dimensions			
	3D	10D	20D	40D
3	50%	100%	100%	100%
4	50%	87.5%	100%	100%
5	50%	80%	100%	100%
6	41.66%	58.33%	91.67%	91.67%
7	57.14%	64.29%	92.86%	78.57%
9	61.11%	55.55%	94.44%	83.33%
10	35%	55%	70%	65%
20	27.5%	50%	55%	52.5%

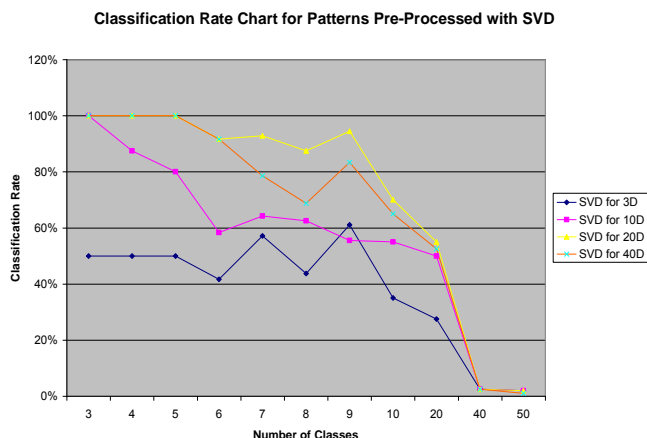


Figure.11 Classification Rate Chart for patterns Pre-Processed with SVD

VI. CONCLUSION AND FUTURE SCOPE

A method for pattern recognition based on Singular Value Decomposition (Feature Extraction) with Error Back Propagation of Neural Network (Recognition) was successfully implemented for iris recognition. Number of experiments was carried out with varying SVD dimensions and number of classes. The results of our work indicate that optimum classification values are obtained with SVD dimensions of 20 and maximum number of classes as 9 with the state-of-the-art computational resources.

For SVD classes around 20 the performance of the network

drops abruptly and becomes independent of SVD vector dimension. This suggests a future scope on researching a better computationally efficient and robust classifier which can handle more number of classes for pattern recognition.

ACKNOWLEDGMENT

The authors wish to acknowledge Institute of Automation, Chinese Academy of Sciences for making CASIA iris image database freely available on web for carrying out research in this field.

REFERENCES

- [1] Li Ma, Tieniu, Tan, "Personal Identification Based on Iris Texture Analysis", IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol 25, No. 10, Dec. 2003, pg 1519-1533
- [2] Daniel Schonberg, darko Kirovski, "Eyeerts", IEEE Trans. On Information Forsensics and Security, Vol 1, No.2, June 2006, Pg 114-153]
- [3] S.Lim, KLee, .O. Byeon, and T.Kim, "Efficient Iris Recognition Through Improvement of Feature vector and Classifier," Journal of ETRI, Vol 23, No.2 pp. 61-70, 2001
- [4] Paulo Eduardo Merlotti, "Experiments on Human Iris Recognition using Error Back-Propagation Artificial Neural Network", Project Report, san Diego State University, April 2004
- [5] Daugman, J., "How Iris Recognition Works", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, Number 1, January 2004.
- [6] Kefeng Fan, Qingqi Pei, Wei Mo, Xinhua Zhao, Qifeng Sun, "An efficient Automatic Iris Image Acquisition and Pre-Processing System", Proc. 2006 IEEE Conf. on Mechatronics and Automation, June 2006, China, pg 1779-1784
- [7] Kittipol Horapong, Jirayut Sreecholpech, Somying Thainimit, Vutipong Areekul, "An Iris Verification using Edge Detection", ID- 07803-9282-5/05@2005 IEEE, pg 1434-1438
- [8] CASIA iris image database, Institute of Automation, Chinese Academy of Sciences, [http://www.sinobiometrics.com]
- [9] Louis L. Scharf, "The SVD and Reduced-Rank Signal Processing in SVD and Signal Processing II: Algorithms, Applications, and Architectures", Pg 3-31, Elsevier Science Publishers, North Holland, 1991.
- [10] Dr. Garcia. E, "Singular Value Decomposition (SVD)- A Fast Track Tutorial", 2006, http:www.misslits.com
- [11] Howard. D, Beale. M, Hagon. M, Neural Network Toolbox for use with MATLAB.

AUTHORS PROFILE



Mr. Babasaheb G. Patil : He received his M.E. Electronics degree in 1990 and B.E. Electronics in 1988. He is currently working as a associate professor in department of Electronics in Walchand College of Engineering, Sangli, Maharashtra, India. He is having keen interest in image processing and communication. He is carrying out research work in the field of Image Processing.



Dr. (Mrs) Shaila Subbaraman : She received M-Tech degree from IISc. Bangalore in 1975 and Ph.D. from IIT Bombay in 1999. She worked in Semiconductor Device Manufacturing company from 1975 to 1989. Currently she is Professor in Department of Electronics in Walchand College of Engineering, Sangli, Maharashtra, India. She has keen interest in the field of Microelectronics and VLSI Design.

Using Semantic Web to support Advanced Web-Based Environment

Khaled M. Fouad

Computer Science Dep., Community College,
Taif Univ., Kingdom of Saudi Arabia (KSA)

Hany M. Harb

Computers and Systems Engineering Dept.,
Faculty of Eng., Al-Azhar Univ., Egypt.

Mostafa A. Nofal

Computer Engineering Dep.,
College of Computer Science and Information Systems,
Taif Univ., Kingdom of Saudi Arabia (KSA).

Nagdy M. Nagdy

Engineering Applications and Computer Systems,
Al-Baha Private College of Science,
Kingdom of Saudi Arabia (KSA)

Abstract—In the learning environments, users would be helpless without the assistance of powerful searching and browsing tools to find their way. Web-based e-learning systems are normally used by a wide variety of learners with different skills, background, preferences, and learning styles.

In this paper, we perform the personalized semantic search and recommendation of learning contents on the learning Web-based environments to enhance the learning environment. Semantic and personalized search of learning content is based on a comparison of the learner profile that is based on learning style, and the learning objects metadata. This approach needs to present both the learner profile and the learning object description as certain data structures. Personalized recommendation of learning objects uses an approach to determine a more suitable relationship between learning objects and learning profiles. Thus, it may advise a learner with most suitable learning objects. Semantic learning objects search is based on the query expansion of the user query and by using the semantic similarity to retrieve semantic matched learning objects.

Keywords- *Semantic Web; Domain Ontology; Learner Profile; Adaptive Learning; Semantic Search ; Recommendation.*

I. INTRODUCTION

Learning environment allows learners to access electronic course contents through the network and study them in virtual classrooms. It brings many benefits in comparison with conventional learning paradigm, e.g. learning can be taken at any time and at any place. However, with the rapid increase of learning content on the Web, it will be time-consuming for learners to find contents they really want to and need to study. The challenge in an information-rich world is not only to make information available to people at any time, at any place, and in any form, but to offer the right thing to the right person in the right way [1].

In the context of e-learning [2], adaptive systems are more specialized and focus on the adaptation of learning content and the presentation of this content. According to [3], an adaptive system focuses on how the profile data is learned by the learner

and pays attention to learning activities, cognitive structures and the context of the learning material.

In Figure 1, the structure of an adaptive system [5] is shown. The system intervenes at three stages during the process of adaptation. It controls the process of collecting data about the user, the process of building up the user model (user modeling) and during the adaptation process.

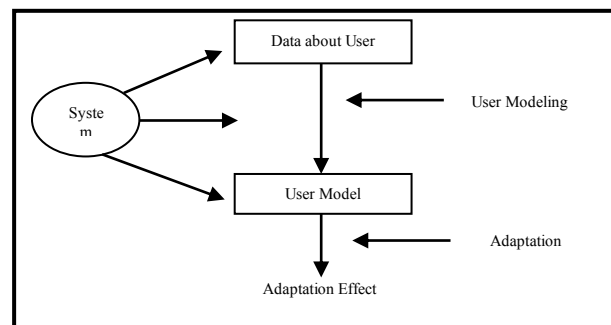


Figure 1: The Structure of an Adaptive System [5]

An advanced e-learning system has to comply with the following requirements [6]:

Personalization: This requirement suggests that the learning process needs to take into account the user's preferences and personal needs. This implies either that the user is in a position to specify explicitly these preferences or that the system has the ability to infer them through a monitoring process. The latter is far more convenient for the end-user and constitutes a highly desirable feature.

Adaptivity: The user's preferences change over time and the system must be able to track them and properly adjust to them. By 'properly', it is implied that the whole history of the user's learning behavior must be taken into consideration, and not just the user's latest (most recent) actions.

Extensibility: An e-learning system has to be extensible in terms of the learning material it provides. The incorporation of new courses and resources must be an easy to accomplish the task.

Interoperability: An e-learning system must be able to both access content from and provide content to digital libraries and other e-learning systems. In this way, the provision of enriched and updated content is feasible.

The semantic web [4] is a space understandable and navigable by both human and software agents. It adds structured meaning and organization to the navigational data of the current web, based on formalized ontologies and controlled vocabularies with semantic links to each other. From the E-Learning perspective, it aids learners in locating, accessing, querying, processing, and assessing learning resources across a distributed heterogeneous network; it also aids instructors in creating, locating, using, reusing, sharing and exchanging learning objects (data and components). The semantic web-based educational systems need to interoperate, collaborate and exchange content or re-use functionality.

Ontology [7] comprises a set of knowledge terms, including the vocabulary, the semantic interconnections, and some simple rules of inference and logic for some particular topic. Ontologies applied to the Web are creating the Semantic Web. Ontologies [8] facilitate knowledge sharing and reuse, i.e. a common understanding of various contents that reach across people and applications. Using ontology in learning environments aims to provide mechanisms to enhance the process of searching and finding learning resources and have the capability to organize and display information that make it easier for learners to draw connections, for instance, by visualizing relationships among concepts and ideas.

Learning environment should not only provide flexible content delivery, but support adaptive content search and recommendation. For better learning experience and effect, search and recommendation of learning content should take into account the contextual information of learners, e.g., prior knowledge, goal, learning style, available learning time, location and interests.

This paper aims to perform the personalized semantic search and recommendation of learning contents on the learning Web-based environments. Semantic and personalized search of learning content is based on a comparison of the learner profile and the learning content description. This approach needs to present both the learner profile and the learning object description as certain data structures. Personalized recommendation of learning objects is based on ontological approach to guide what learning contents a learner should study, i.e. what learning objects a course should have according to learner preference and intention.

II. RELATED WORKS

Personalized search [9] is addressed by a number of systems. Persona [10] uses explicit relevant feedback to update user profiles that are represented by means of weighted open directory project taxonomy [11]. These profiles are used to filter search results. Personalized variants of PageRank, as found in Personalized Google or the Outride Personalized Search System [12]. Authors in [13] re-rank the search results of queries for medical articles profiles keywords, associated concepts, and weights generated from an electronic patient

record. In [14], it was filtered search results on the grounds of user profiles obtained from earlier queries. These profiles consist of a set of categories, and weighted terms associated with each category. In their work on personalizing search results, [15] they distinguish between long-term and short-term interests. While aiming at personalization in a broader sense, [16] use click-through data to increase the performance of search results.

In the paper [17], authors have proposed an approach to personalized query expansion based on a semantic user model. They discussed the representation and construction of the user model which represents individual user's interests by semantic mining from user's resource searching process, in order to perceive the semantic relationships between user's interests which are barely considered in traditional user models and to satisfy the requirement of providing personalized service to users in e-Learning systems. They exploited the user model to provide semantic query expansion service in our e-Learning system.

Authors in [18] have shown that extracting the semantic interests of learner profiles can form a reasonable and simple way to represent the learning context, and that semantic learner profile, coupled with a semantic domain ontology that represents the learned content, enhance the retrieval results on a real e-learning platform.

This paper [19] proposed a new method for the personalized search, using click-through data as the personal data. Firstly, uses the semantic statistical of word frequency method to extract the query expansion terms and recommended to the user. Secondly, improves the Naive Bayesian classifier and combines SVM to make users' personalized learning models, then provides personalized re-sort results by user models. After experimental evaluation, it showed that this method has a significant effect, not only provides a meaningful query expansion terms, but also significantly improves the ranking of results.

The study in [20] authors proposed an ontological approach for semantic-aware learning object retrieval. The proposed ontological approach has two significant novelties: a fully automatic ontology query expansion algorithm for inferring and aggregating user intentions based on their short queries.

This paper [21] proposed a personalized e-learning method based on hybrid filtering. Two-level user profiles direct the recommendation process. Group profile reflects the users whose similar learning needs are similar with the current user. Topic profile describes the user's interests with topics that the user has learned. Group profile and topic profile are bases of collaborative filtering recommendation and content-based filtering recommendation respectively.

In the paper [22], the authors introduced the principle and implementation steps of Collaborative Filtering (CF) algorithm. Then a novel CF recommendation algorithm was proposed on the combination of user profile weight and time weight. In this way, on one hand, the improved prediction can discover user's latent demands more precisely. On the other hand, it also can sense the changes of user's preference and then adjust the recommendation promptly.

III. THE PROPOSED SYSTEM

Personalized service is being paid close attention as a new method of intelligent information service to satisfy the increasing informational demands of the users in different systems. User model plays an important role in providing personalized service by representing the user's identity information and interests. There are many user models which have been adopted in various systems to acquire interests of users. In the e-Learning scenario, the learner model is exploited to represent the interests and background knowledge of individual learners [23]. The key technology of providing personalized learning services is to represent and acquire user's interests that are used in user modeling. User modeling is used to search and recommend content relevant to user interests.

In our proposed approach, personalized search of learning objects in e-learning is based on a comparison of the learner profile and the learning object (resource) description [24, 25]. Because such an approach needs to present both the learner profile and the learning object description as certain data structures, it requires the development of ontological models [26, 9] of the learner and learning object.

The proposed approach has two aspects, first for personalized search of learning objects is generally described in [24, 27], second for personalized recommendation suitable learning objects is proposed in [28, 29].

The key idea of the Semantic Web is to have data defined and linked in such a way that its meaning is explicitly interpretable by software processes rather than just being implicitly interpretable by humans. The Semantic Web can represent knowledge, including defining ontologies as metadata of resources. Ontology is a formal, semantic specification of a conceptualization of a domain of interest. Ontologies are used to describe the semantics of information exchange.

The metadata used in our work, data about data, is to provide structured information that describes, locates and

explains information resources making it easier for resources to be retrieved. It is important to remember that data and metadata are different. Data is values, individual parts of information, whereas metadata describes the relationship between the parts and other data. Together data and metadata make information portable, because the relationships among the data values remain separate from their storage. Metadata is a key concept in developing the Semantic Web, to allow computers to share information automatically, data and metadata must be grouped together. Therefore, to ensure metadata can be automatically processed by machines, some metadata standard is needed [46].

The learner model is abstract expression to the learner characteristic. The learner model is not the expression of learner's all characteristics, but to describe and express partial learner characteristics according to the different learning system's needs.

The present research will describe details of building the learner and learning object ontological models to perform personalized search in learning objects and recommendation suitable learning objects to learners.

In order to implement the proposed personalized search of learning objects according to the created ontological models of the learner and learning object, some IMS Learner Information Package Specification corresponding to some IEEE LOM [30] standard have been chosen, and the criteria to estimate conformity of LOM to the learner personal profile with the coefficients of importance. Our proposed system architecture is shown in figure 2.

Our system aims to perform these objectives:

1. Presenting a technical solution to an approach and methodology for personalized search of learning objects according to criteria that determine the learner's interests.
2. Proposing an approach to adjust a learner's interests. Because different attributes have different importance

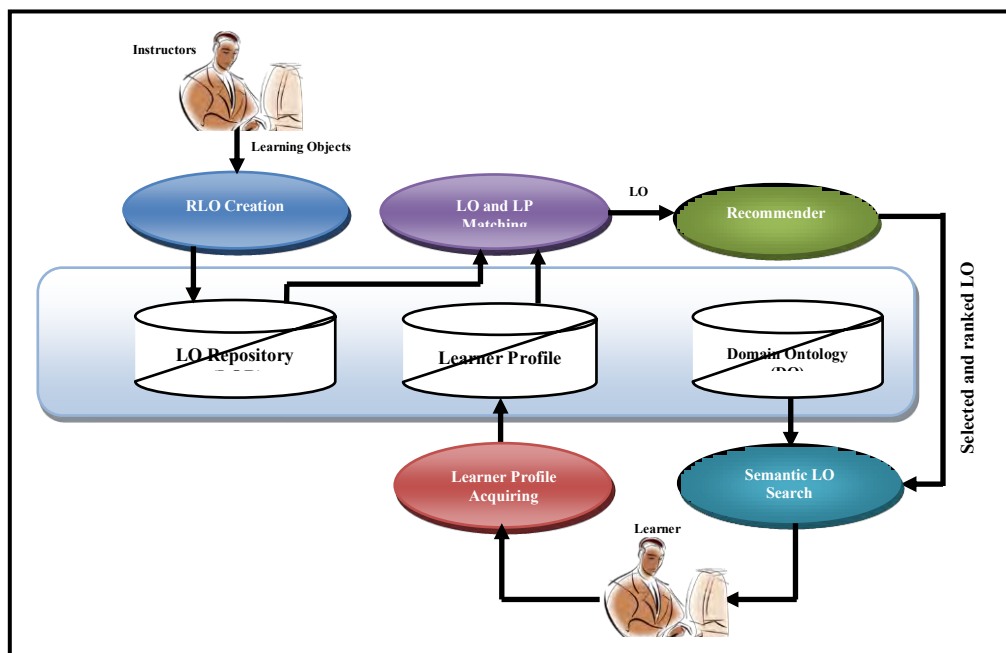


Figure 2: The Proposed System Architecture

to different learners, the system adjusts the weights in comparison to the learner's interests and goals.

3. Proposing personalized learning objects search in E-learning, which is intended to allow a learner (user) to create a learner profile describing his/her personal interests using the IMS attributes [31], and then run a personalized search in a learning object repository on the Web to find relevant learning objects, which match that learner profile.
4. Ranking available learning objects by comparing values of corresponding attributes in the learner profile and learning object metadata (LOM).
5. Using specific ontology to infer what learning objects are needed for a course established for a specific learner requiring a specific subject and how to look for them on the Internet.
6. Recommending suitable learning objects according to a user's preference and intention.
7. Referring to the experiences of similar users and adopting neighbor-interest to look for the learning objects that the user should be interested.
8. Providing adaptive, personalized recommendation for each user (learner).

A. Reusable Learning Objects (RLO) Creating

As mentioned in [32], authors have determined that the reusable learning objects is a reusable chunk of content with the following two fundamental properties: first is instructional sound content with the focused learning objectives. Second property is the facility that allows the learner to practice, learn, and receive assessment. Also, they define the sharable learning objects as RLO with the additional interoperability property that is the metadata or keywords that describe the object's attributes and mechanisms for communicating with any e-learning system. The aim of this methodology is to select and extract as much of the existing raw content into RLO. The methodology is an iterative five step process to select appropriate content for the RLO with opportunities to refine and re-structure as the extraction is taking place. The algorithm for building the RLO is shown in figure 3. A learning object must be modular, discoverable and interoperable, in order to be reused. To achieve these features and improve efficiency many people have dedicated long hours of hard work.

Input: <i>Learning Material</i>
Output: <i>Reusable learning objects of the learning material.</i>
Procedure:
1. Create detailed table of contents for the material.
2. Define set of learning objectives for some of the topic/subtopic.
3. Select raw content to achieve each identified learning objective.
4. Include the review Question/Answer.
5. Include the examination Question/Answer.

Figure 3: Algorithm for building the RLO

The majority of the efforts focus on the definition of standardization. Organizations such as IEEE [30] have contributed significantly by defining indexing standards called metadata (data about data). Metadata structures [33] contain information to explain what the leaning object is about, how to search, access, and identify it and how to retrieve educational content according to a specific demand.

The IEEE LOM standard specification specifies a standard for learning object metadata. It specifies a conceptual data schema that defines the structure of a metadata instance for a learning object. The IEEE LOM specification consists of nine categories, which includes 60 data elements. Each category has a specific purpose, such as describing general attributes of objects, and educational objectives. Table 1 shows the LOM categories adopted in our work.

TABLE I. THE MAIN CATEGORIES OF IEEE LOM

Category Name	Category Fields	Description
General	Identifier, Catalog, Entry, Title, Language, Description, Keyword, Coverage, Structure, Aggregation Level	general information that describes the learning object as a whole.
Technical	Format, Size, Location, Requirement, OrComposite, Type, Name, Minimum Version, Maximum Version, Installation Remarks, Other Platform Requirements, Duration	technical requirements and characteristics of the learning object.
Educational	Interactivity Type, Learning Resource Type, Interactivity Level, Semantic Density, Intended End User Role, Context, Typical Age Range, Difficulty, Typical Learning Time, Description, Language,	key educational or pedagogic characteristics of the learning object.

One of the chartered activities of the IEEE LTSC is to develop an XML binding for LOM [34]. This activity is ongoing, but the standard XML binding has not yet been approved and published. While the LOM standard defines the structure of a metadata instance, it does not define how a learning technology system will represent or use a metadata instance for a learning object.

The XML Binding defines an exchange format for metadata. With XML, course developers may put semi-structured information, such as the course content or course structure, into a discrete relational field, and then work with this information as with structured blocks of data, not as with a string of bytes. In our research, we describe each Learning Object by means of the XML document validated against an XML Schema defined by the IEEE LOM standard. Figure 4 shows the used LOM category and its fields in our system as relationship diagram of database.

We choose the tags from the standard schema, so every tag in our schema is still meaningful to others. A third-party search engine that can handle the XML metadata documents conforming to the standard schema could also handle ours. Figure 5 shows the part of schema for learning objects metadata (generated by XML Editor [35]) and Figure 6 shows the part of DTD of the XML file of learning objects metadata.

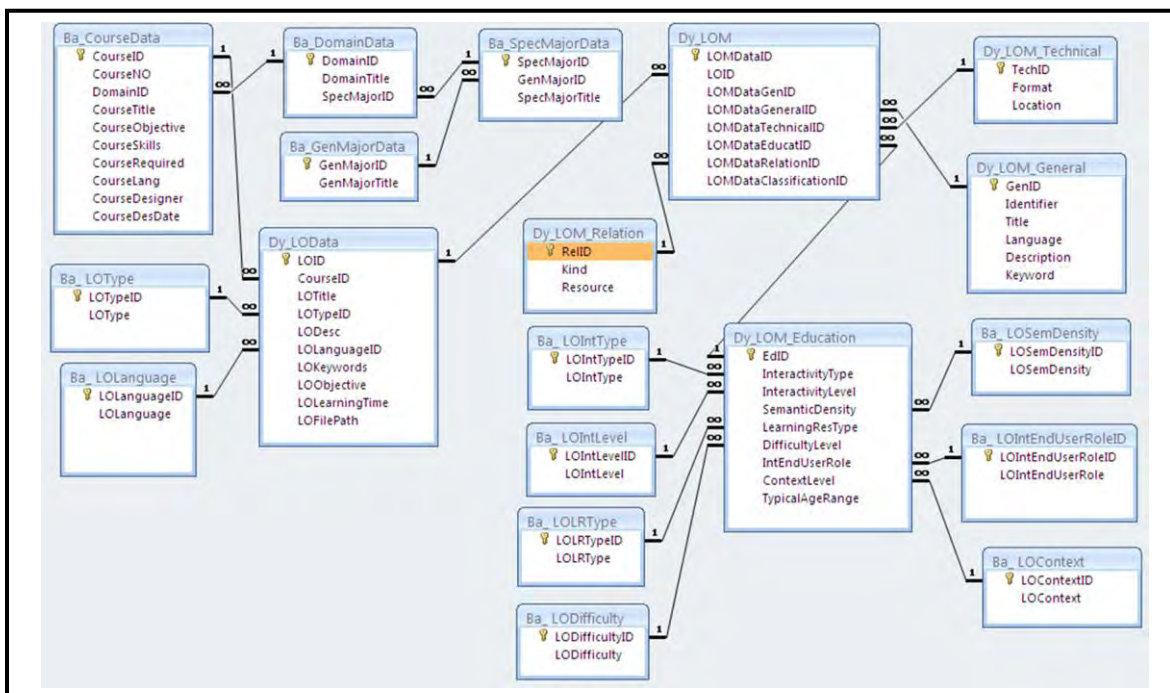


Figure 4: The used LOM category and its fields in our system as relationship diagram of database

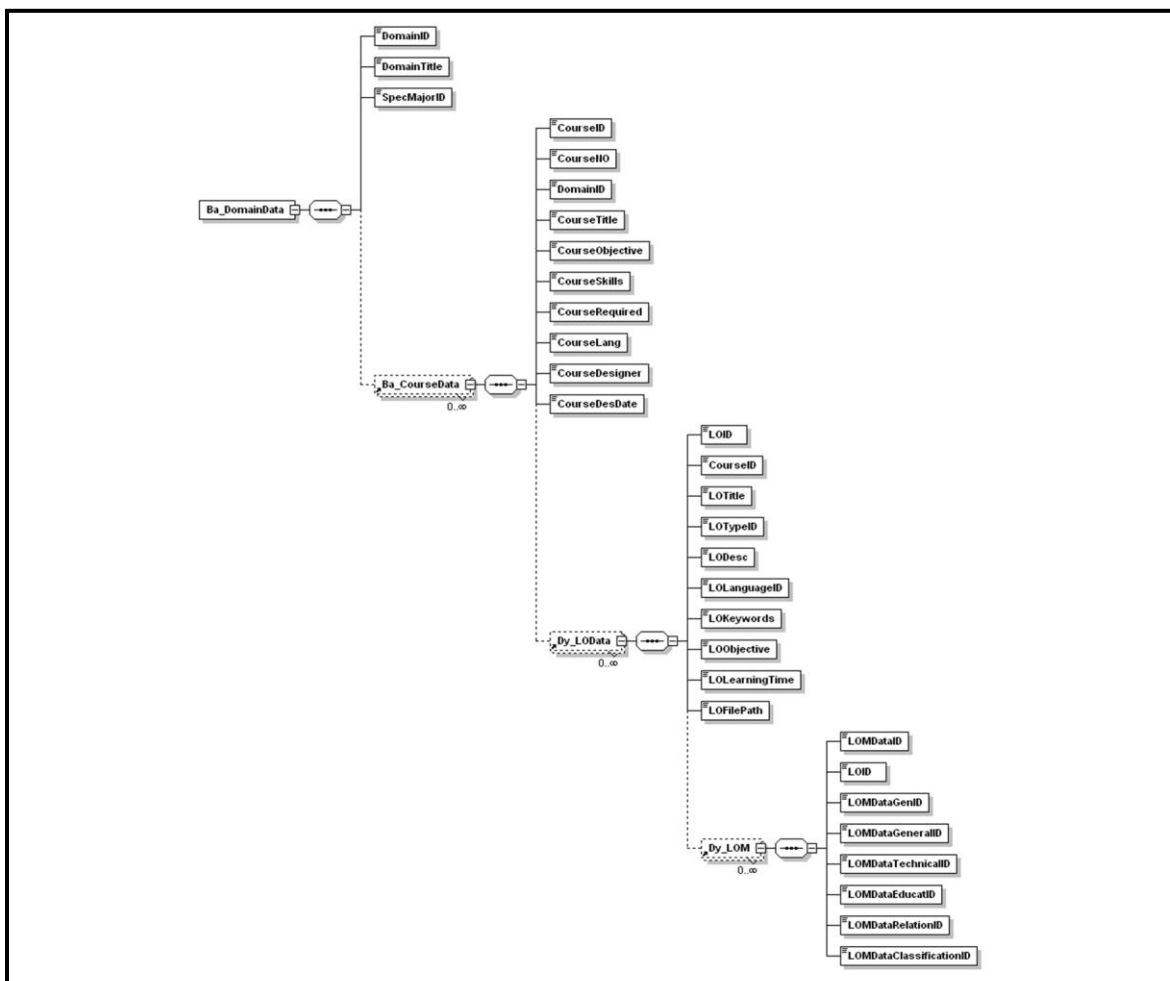


Figure 5: The part of schema for learning objects metadata

```
<?xml version="1.0" encoding="UTF-8"?>
<ELEMENT Ba_LOContext (LOContextID, LOContext, Dy_LOM_Education*)>
<ELEMENT Ba_LODifficulty (LODifficultyID, LODifficulty, Dy_LOM_Education*)>
<ELEMENT Ba_LOIntEndUserRole (LOIntEndUserRoleID, LOIntEndUserRole,
Dy_LOM_Education*)>
<ELEMENT Ba_LOIntLevel (LOIntLevelID, LOIntLevel, Dy_LOM_Education*)>
<ELEMENT Ba_LOIntType (LOIntTypeID, LOIntType, Dy_LOM_Education*)>
<ELEMENT Ba_LOLRTType (LOLRTTypeID, LOLRTType, Dy_LOM_Education*)>
<ELEMENT Ba_LOLanguage (LOLanguageID, LOLanguage, Dy_LOData*)>
<ELEMENT Ba_LOSemDensity (LOSemDensityID, LOSemDensity, Dy_LOM_Education*)>
<ELEMENT Ba_LOType (LOTypeID, LOType, Dy_LOData*)>
<ELEMENT Ba_CourseData (CourseID, CourseNO, DomainID, CourseTitle, CourseObjective,
CourseSkills, CourseRequired, CourseLang, CourseDesigner, CourseDesDate, Dy_LOData*)>
<ELEMENT Ba_DomainData (DomainID, DomainTitle, SpecMajorID, Ba_CourseData*)>
<ELEMENT Ba_GenMajorData (GenMajorID, GenMajorTitle, Ba_SpecMajorData*)>
<ELEMENT Ba_SpecMajorData (SpecMajorID, GenMajorID, SpecMajorTitle, Ba_DomainData*)>
<ELEMENT Dy_LOData (LOID, CourseID, LOTitle, LOTypeID, LODesc, LOLanguageID,
LOKeywords, LOObjective, LOLearningTime, LOFilePath, Dy_LOM*)>
<ELEMENT Dy_LOM (LOMDataID, LOID, LOMDataGenID, LOMDataGeneralID,
LOMDataTechnicalID, LOMDataEducID, LOMDataRelationID, LOMDataClassificationID)>
<ELEMENT Dy_LOM_Education (EdID, InteractivityType, InteractivityLevel, SemanticDensity,
LearningResType, DifficultyLevel, IntEndUserRole, ContextLevel, TypicalAgeRange, Dy_LOM*)>
<ELEMENT Dy_LOM_General (GenID, Identifier, Title, Language, Description, Keyword,
Dy_LOM*)>
<ELEMENT Dy_LOM_Relation (RelID, Kind, Resource, Dy_LOM*)>
<ELEMENT Dy_LOM_Technical (TechID, Format, Location, Dy_LOM*)>
<ELEMENT CourseTitle (#PCDATA)>
<ELEMENT CourseObjective (#PCDATA)>
<ELEMENT CourseSkills (#PCDATA)>
<ELEMENT CourseRequired (#PCDATA)>
<ELEMENT CourseLang (#PCDATA)>
<ELEMENT CourseDesigner (#PCDATA)>
<ELEMENT CourseDesDate (#PCDATA)>
<ELEMENT DomainID (#PCDATA)>
<ELEMENT DomainTitle (#PCDATA)>
<ELEMENT SpecMajorID (#PCDATA)>
<ELEMENT GenMajorID (#PCDATA)>
<ELEMENT GenMajorTitle (#PCDATA)>
<ELEMENT SpecMajorID (#PCDATA)>
<ELEMENT GenMajorID (#PCDATA)>
<ELEMENT SpecMajorTitle (#PCDATA)>
<ELEMENT LOID (#PCDATA)>
<ELEMENT CourseID (#PCDATA)>
<ELEMENT LOTitle (#PCDATA)>
<ELEMENT LOTypeID (#PCDATA)>
<ELEMENT SemanticDensity (#PCDATA)>
<ELEMENT LearningResType (#PCDATA)>
<ELEMENT DifficultyLevel (#PCDATA)>
<ELEMENT IntEndUserRole (#PCDATA)>
<ELEMENT ContextLevel (#PCDATA)>
<ELEMENT TypicalAgeRange (#PCDATA)>
<ELEMENT GenID (#PCDATA)>
<ELEMENT Identifier (#PCDATA)>
<ELEMENT Title (#PCDATA)>
<ELEMENT Language (#PCDATA)>
<ELEMENT Description (#PCDATA)>
<ELEMENT Keyword (#PCDATA)>
<ELEMENT RelID (#PCDATA)>
<ELEMENT Kind (#PCDATA)>
<ELEMENT Resource (#PCDATA)>
<ELEMENT TechID (#PCDATA)>
<ELEMENT Format (#PCDATA)>
<ELEMENT Location (#PCDATA)>
```

Figure 6 The part of DTD of the XML file of learning objects metadata

TABLE II. CATEGORIES OF FELDER-SILVERMAN’S LEARNING STYLE

Learning Style Category	Description
Sensing vs. Intuitive	It represents the abstraction level of the learning material the learner prefers. A sensing learner likes learning facts and needs more practical case studies. An intuitive learner usually prefers innovation and dislikes repetition.
Visual vs. Verbal	It indicates whether the learner prefers auditory (textual) or visual documents.
Active vs. Reflective	It indicates how the learner prefers to process information: actively (through engagement in activities or discussions) or reflectively (through introspection)
Sequential vs. Global	It indicates how the learner progresses toward understanding. Sequential learners prefer sequential explanations while global learners usually prefer an initial overview of the involved topics which possibly shows them the most important steps and relations they are going to study

B. Learner Profile Acquiring using learning style

There are five popular and useful features when is viewing the learner as an individual, these are: the learner’s knowledge, interests, goals, background, and individual traits [36]. Learning styles are typically defined as the way people prefer to learn. We can represent the learning style in stereotype model according to the Felder-Silverman’s learning style categories. From the perception, input processing and understanding four dimensions, the Felder-Silverman’s learning style categories are shown in table 2 [37, 38].

The learner actions that can be used to identify learner cognitive traits in learning systems by learner behaviors that can enable to acquire the learning style. Number of these actions is shown in [39]. Example of the actions that can enable to acquire learning styles base on Felder-Silverman model (FSLSM) is found in table 3.

TABLE III. THE RELATIONSHIP BETWEEN LEARNER ACTIONS AND (FSLSM) CATEGORY

Parameter	Value	FSLSM Category
No. of visits/postings in forum/chat	High	Active, Verbal
No. of visits and time spent on exercises	High	Active, Intuitive
Amount of time dealt with reading material	High	Reflective
Performance on questions regarding theories	High	Intuitive
Performance on questions regarding facts	High	Sensing
Amount of time spent on a Test	High	Sensing
No. of revisions before handing in a test	High	Sensing
No. of performed tests	High	Sensing
No. of visits and time spent on examples	High	Sensing
Amount of time spent on contents with graphics	High	Visual
Performance in questions related to graphics	High	Visual
Performance on questions related to overview of concepts and connections between concepts	High	Global
Performance on questions related to details	High	Sequential
Performance on tests in General	High	Sequential
No. of visits and time spent on outlines	High	Global
Navigation pattern	Skipping learning objects	Global
Navigation pattern	Linear	Sequential

Another action that is found in [40] as the number of rules to describe learner learning style by recording the learner behavior in the system as found in figure 7.

```

IF learner does not know the answer;
THEN
Show learner image/diagram;
-----
IF learner shown image/diagram AND learner gives correct answer;
THEN
Increase VISUAL;
-----
IF answer is given in the explanation text AND learner does not
know the answer;
THEN
Increase INTUITOR AND Increase VISUAL;
    
```

Figure 7 Example of rules used to adjust learner learning style

C. Learning Content Recommendation and Matching

Personalized recommendation is a widely used application of Web personalized services which alleviate the burden of information overload by collecting information which meets the user's needs. An essential of Web recommendation is how to build user profile, which involves the information and preference of user and has a great impact on the performance of Web personalized recommendation. The Adaptive Systems and Recommender Systems [41] are focused in exploring a certain hypermedia structure in order to help user finding the best way for their interests, while the Recommender Systems are focused on a network of Web resources, bind by existing or virtual relations, aiming to provide users with individual views on Web data.

The Felder-Silverman Learning Style Model is described by the dimensions of Learning and Teaching Styles [42], creating a relationship to learning styles and teaching strategies that could be adopted to support the learner learning style [43].

Zaina and Bressan in [44, 45] proposed an alternative approach that splits the learner learning profile (preferences) into three categories: perception, presentation format and learner participation. Along the text, this altered model is referred to as preference categories; its goal is to detect clusters of preferences that reflect different data perspectives caught during the tracking of learning styles.

Each category has a teaching-method correspondence that defines the matching with the learners' learning styles, as predicted in the Felder/Silverman proposal as found in [45]. According to Felder and Silverman, the teaching-learning style corresponds to the values of LOM category fields. The Example to show the relationship between LOM educational fields and the preferences category is shown in table 4.

By matching the learning objects metadata, that is stored in learning objects repository, with the learner profile in the system, the system can recommend the learning objects based on the learning styles.

D. The Domain Ontology

The main reason for ontology [47] is to enable communication between computer systems in a way that is independent of the individual system technologies, information architectures and application domain.

TABLE IV. THE RELATIONSHIP BETWEEN LOM EDUCATIONAL FIELDS AND THE PREFERENCES CATEGORY

Preference Categories	Features	Learning Styles	Teaching Methods	LOM – Educational Field	LOM – Educational Field Value
Perception	The focus is in the best way through which the learner can obtain information: contents, exercise types, for instance.	Sensing	Concrete	Interactivity	Active
		Intuitive	Abstract		Expositive
Presentation Format	It is related to the input. Content preferences chosen by the learner such as media types.	Visual	Visual	Learning Resource Type	Figure, Video, Film, and others
		Auditory	Verbal		Text, Sound, and Format others
Learner Participation	It represents the learner preferences for the activities participation or observation.	Active	Active		Practical Exercise, Experiment, and others
		Reflective	Passive		Questionnaire and Readings

Ontology includes rich relationships between terms and each specific knowledge domain and organization will structure its own ontology which will be organized into mapped ontology. The domain of our learning content and the ontology we have developed within proposed system is that of computer science. The ontology covers topics like artificial intelligence, communications; computational theory, computer graphics, data structures, database, programming, etc. It is used mainly to index the relevant learning objects and to facilitate semantic search and re-usability of learning objects.

It was proposed in [48] a knowledge engineering approach to build domain ontology. Figure 8 shows main steps of the ontology development process.

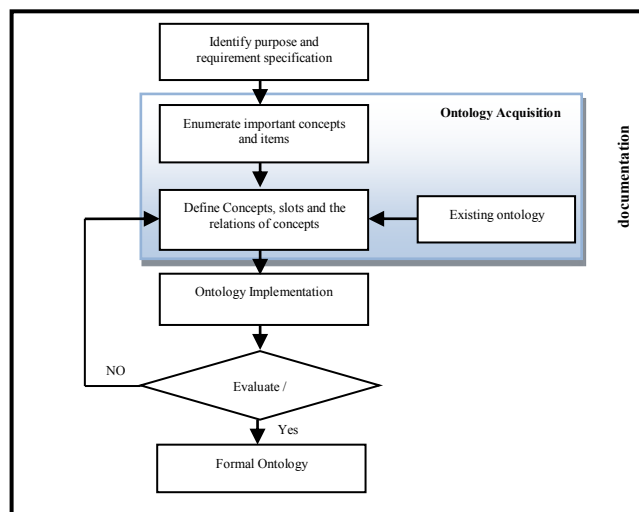


Figure 8 Main steps of the ontology development process

We use protégée [49, 50] as our ontology tool. Since protégée is an open source ontology editor, developed by Stanford Center for Biomedical Informatics Research and coded by JAVA. Protégé interface style is similar to Windows applications' general style, so it is easy to learn and use. Figure 9 shows part of our domain ontology and the extracted.

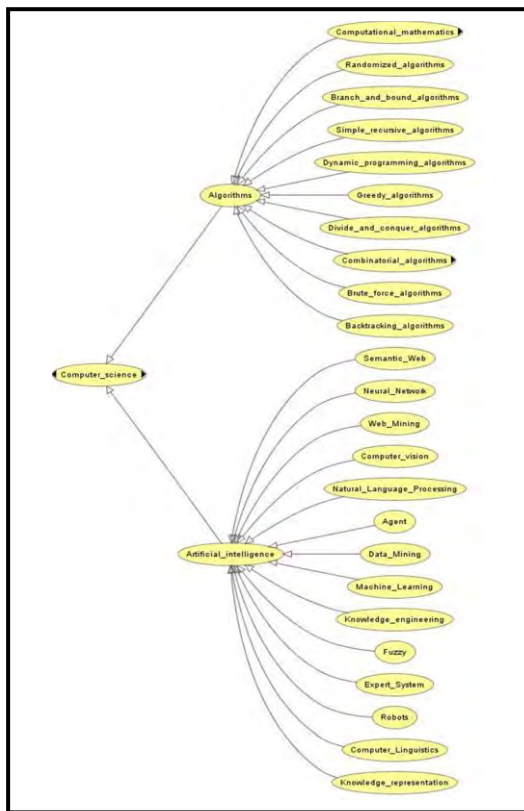


Figure 9 Part of our domain ontology

E. Semantic LO Search

According to keyword-based search present serious problems related to the quality of the search results. It often happens that relevant pages are not indexed by a traditional search; in this case important information can be reached only if its specific internet address is known. Moreover, searches based on keywords are very closely related to the spelling of the word and not to its meaning. One current problem of information search issues is that it is not really possible to automatically extract meaning from the relevant results of a query. One main reason for this is that the web was initially designed for direct human use and thus the documents do not provide machine readable semantic annotations. This work focuses on the first of these items, specifically in the formulation and the user's query processing. We expect to prove that through linguistic processing, the use of dictionaries and domain ontologies, the instructional designer's query terms become more specific.

The Semantic Search process includes the steps, that is appeared in the next algorithm, is shown in the figure 10.

CONCLUSION

The adaptive learning system provides support to the learner according to the individual characteristic. It can provide a learner view adapt to learner personalization characteristic, which not only includes personalized resources, but also includes the personalized learning process and strategy. So we should establish a learner model for each learner, containing the information such as state-of-art of learner, the goal and

interest and so on. The system reduces the information spaces for learner browsing according to the learner model in the application, and presents the most interesting information to the learner.

Input: Query of User.
Output: Retrieved Semantic Information
Procedure:

1. Tokenizing query keywords to number of terms.
2. Remove the stop words.
3. Stem the word.
4. Get POS (Part of Speech) of the word in the query.
5. Expand the words by the hypernym and hyponym concepts in the Wordnet.
6. Expand the words by the Domain Ontology (DO) as:
 - a. Search the word in the DO.
 - b. Check if the word is the root or not.
 - i. If Yes
 1. Get the Hyponym, and Get the neighbor node.
 2. Add the two concepts to the expanded query.
 - ii. If No and is not the Leaf.
 1. Get the Hyponyms Hypernyms, and neighbor.
 2. Add the two concepts to the expanded query.
 - iii. If No and is the Leaf.
 1. Get *Hyponyms*, and neighbor.
 2. Add the two concepts to the expanded query.
 - c. Compute the similarity between concepts, put N pre-expansion words that has high relativity as expansion words.
 - d. Add Expanded Query to the original query.
7. Use Semantic Similarity between the expanded words and the terms in the LO
8. Rank the LO based on the high semantic similarity weight.
9. Return the ranked LO from LOR.

Figure 10 Algorithm of Semantic Search of LO

This work presented a technical solution to an approach and methodology for personalized search and recommendation of learning objects according to the learner's profile. Adaptive recommendation model is to retrieve and recommend for a learner suitable learning objects.

In this work, we have defined a methodology that links learning objects metadata and learning profiles for automatic content recommendation. To do so, we have used the Felder-Silverman Learning Style Model along with the IEEE LOM standard, a combination that, extending former works, can suitably relate learner profiles and learning objects, automatically, in different fields of learning, and consistently reflecting the intrinsic style of the learners.

The semantic search of the learning objects is based query expansion and using semantic similarity between the learning objects and the query keywords.

REFERENCES

- [1] Y. Zhiwen, N. Yuichi, J. Seie, K. Shoji, and M. Kenji. (2007), Ontology-Based Semantic Recommendation for Context-Aware E-Learning, UIC 2007, LNCS 4611, pp. 898–907, 2007, Springer-Verlag Berlin Heidelberg.
- [2] F. Christoph, (2005). User Modeling and User Profiling in Adaptive E-learning Systems, Master's Thesis At Graz University of Technology.
- [3] [3] Felix M'odritscher, (2004). Victor Manuel Garcia-Barrios, and Christian Gutl. The Past, the Present and the future of adaptive E-Learning. In Proceedings of the International Conference Interactive Computer Aided Learning (ICL2004), 2004. http://www.iicm.edu/iicm_papers/icl2004/adaptive_e-learning/adaptiv_e-learning.pdf.
- [4] D. Dicheva. (2008), Ontologies and Semantic Web for E-Learning, In : "Handbook on Information Technologies for Education and Training", 978-3-540-74155-8, Springer Berlin Heidelberg.
- [5] B. Peter and T. Mark. (2002) From Adaptive Hypermedia to the Adaptive Web. Communications of the ACM, Volume 45 Issue 5.
- [6] K. Yiouli, D. Panagiotis, A. Evgenia, D. Konstantinos, T. Michael, P. Maria. (2008). User Profile Modeling in the context of web-based learning management systems. Journal of Network and Computer Applications 31 (2008) 603–627. Elsevier Ltd.
- [7] G. Fayed, D. Sameh, H. Ahmad, M. Jihad., A. Samir, and S. Hosam. (2006). E-Learning Model Based On Semantic Web Technology, International Journal of Computing & Information Sciences Vol. 4, No. 2, August 2006, On-Line. Pages 63 – 71.
- [8] Z. Thomas, and V. Juan. (2010). Towards an Ontology for the Description of Learning Resources on Disaster Risk Reduction. WSKS 2010, Part I, CCIS 111, pp. 60–74, Springer-Verlag Berlin Heidelberg.
- [9] G. Susan, S. Mirco, and P. Alexander. (2007), Ontology-Based User Profiles for Personalized Search, DOI10.1007/978-0-387-37022-4, Springer US.
- [10] F. Tanudjaja, L. Mui. (2002) Persona: A Contextualized and Personalized Web Search. Proc 35 th Hawaii Intl. Conf. on System Sciences.
- [11] The Open Directory Project (ODP). <http://dmoz.org>.
- [12] J. Pitkow, H. Schütze, T. Cass et all. (2002), Personalized search. CACM 2002; 45(9):50-55.
- [13] K. McKeown, N. Elhadad, V. Hatzivassiloglou.(2003), Leveraging a common representation for personalized search and summarization in a medical digital library. In Proceedings of the 3 rd ACM/IEEE-CS joint conference on Digital libraries 2003; 159-170.
- [14] F. Liu, C. Yu, W. Meng. (2002) Personalized web search by mapping user queries to categories. In Proceedings CIKM'02 2002; 558-565.
- [15] K. Sugiyama, K. Hatano, M. Yoshikawa. (2004), Adaptive web search based on user profile constructed without any effort from users. In Proceedings 13 th Intl. Conf. on World Wide Web 2004; 675-684.
- [16] R. Almeida, V. Almeida. (2004), A Community-Aware Search Engine. In Proceedings of the 13 th International Conference on the World Wide Web, May 2004.
- [17] L. Xiaojian, C. Shihong. (2009). Personalized Query Expansion Based on Semantic User Model in e-Learning System. 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery. 978-0-7695-3735-1/09, IEEE.
- [18] Z. Leyla, N. Olfa. (2008). Semantic Information Retrieval for Personalized E-learning. 2008 20th IEEE International Conference on Tools with Artificial Intelligence. 1082-3409/08, IEEE.
- [19] C. Cheqian, L. Kequan, L. Heshan, D. Shoubin. (2010). PERSONALIZED SEARCH BASED ON LEARNING USER CLICK HISTORY. Proc. 9th IEEE Int. Conf. on Cognitive Informatics (ICCI'10). 978-1-4244-8040-1/10, IEEE.
- [20] L. Ming-Che, T. Kun, W. Tzone. (2008). A practical ontology query expansion algorithm for semantic-aware learning objects retrieval. Computers & Education 50 (2008) 1240–1257. Elsevier Ltd.
- [21] D. Lianhong, L. Bingwu, T. Qi. (2010). Hybrid Filtering Recommendation in E-Learning Environment. 2010 Second International Workshop on Education Technology and Computer Science. 978-0-7695-3987-4/10, IEEE.
- [22] D. Chen, D. Xu. (2009). A Collaborative Filtering Recommendation Based on User profile weight and Time weight. 978-1-4244-4507-3/09, IEEE.
- [23] L. Xiaojian, C. Shihong. (2009), Research on Personalized User Model Based on Semantic Mining from Educational Resources Searching Process, 2009 International Joint Conference on Artificial Intelligence, 978-0-7695-3615-6/09, IEEE.
- [24] B. Yevgen, B. Hamidreza, K. Igor, F. Michael. (2009), An adjustable personalization of search and delivery of learning objects to learners, Expert Systems with Applications 36 (2009) 9113–9120, doi:10.1016/j.eswa.2008.12.038, Elsevier Ltd.
- [25] K. Igor, L. Natalya, M. Sergiy, T. Vagan.(2004), Personalized Distance Learning Based on Multiagent Ontological System, Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT'04), 0-7695-2181-9/04, IEEE .
- [26] Z. Hui, S. Yu, and S. Han-tao. (2007). Construction of Ontology-Based User Model for Web Personalization, C. Conati, K. McCoy, and G. Paliouras (Eds.): UM 2007, LNAI 4511, pp. 67–76, Springer-Verlag Berlin Heidelberg.
- [27] K. Igor, R. Victoria, B. Yevgen. (2006), Building Learner's Ontologies to Assist Personalized Search of Learning Objects, ICEC'06, August 14–16, 2006, Fredericton, Canada, Copyright 2006 ACM 1-59593-392-1.
- [28] W. Tzone, H. Kun, L. Ming and C. Ti Kai. (2007), Personalized Learning Objects Recommendation based on the Semantic-Aware Discovery and the Learner Preference Pattern, Educational Technology & Society, 10 (3), 84-105.
- [29] T. Kun, C. Ti Kai, L. Ming Che, W. Tzone. (2006), A Learning Objects Recommendation Model based on the Preference and Ontological Approaches, Proceedings of the Sixth International Conference on Advanced Learning Technologies (ICALT'06), 0-7695-2632-2/06, IEEE.
- [30] Draft Standard for Learning Object Metadata (LOM). IEEE P1484.12.1, IEEE Learning Technology Standards Committee (2004), http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf.
- [31] <http://www.imsproject.org>.
- [32] S. Rajendra, B. Margaret, G. Ross. (2004). Creating sharable learning objects From Existing Digital Course Content. WCAE '04 Proceedings of the 2004 workshop on Computer architecture education. ACM.
- [33] A. Prasad and P. Devika. (2009). Metadata for Resource Discovery in Learning Repositories Road to enhanced access to e-learning content. International Workshop on Technology for Education (T4E), Aug 4-6, 2009, Bangalore. 978-1-4244-5505-8/09, IEEE.
- [34] M. Pasquale, G. Alfredo, T. Giorgio, U. Domenico. (2007). Personalizing learning programs with X-Learn, an XML-based, "user-device" adaptive multi-agent system. Information Sciences 177 (2007) 1729–1770. Elsevier Inc.
- [35] Altova XMLSpy® 2012 Enterprise Edition, http://www.altova.com/download/xmlspy/xml_editor_enterprise.html.
- [36] B. Peter, M. Eva. (2007). User Models for Adaptive Hypermedia and Adaptive Educational Systems. The Adaptive Web, LNCS 4321, pp. 3 – 53, Springer-Verlag Berlin Heidelberg.
- [37] Enver Sangineto. (2008). An Adaptive E-Learning Platform for Personalized Course Generation". In Claus Pahl(ed) Architecture Solutions for E-Learning Systems. IGI Publishing.
- [38] C. Shipin, Z. Jianping. (2008). The Adaptive Learning System based on Learning Style and Cognitive State. 2008 International Symposium on Knowledge Acquisition and Modeling. 978-0-7695-3488-6/08, IEEE.
- [39] P. Elvira, T. Philippe and B. Costin. (2007). Adaptive Educational Hypermedia Systems: A Focus on Learning Styles. EUROCON 2007 The International Conference on "Computer as a Tool". 1-4244-0813-X/07, IEEE.
- [40] M. Annabel, A. Keeley, A. David, E. Bruce, and S. Karen. (2010). Oscar: An Intelligent Conversational Agent Tutor to Estimate Learning Styles. 978-1-4244-8126-2/10, IEEE.

- [41] B. Mihaela, S. Florence, Z. Corinne. (2009). Ontology-Based User Competencies Modeling for E-Learning Recommender Systems. IGI Global.
- [42] F. RICHARD, B. REBECCA. (2005). Understanding Student Differences”, Journal of Engineering Education, 94, 1, 2005, pp 57-72.
- [43] F. Richard, L. Silverman. (1988) “Learning and Teaching Styles in Engineering Education”, Journal of Engineering Education, 78, 7, 1988, pp. 674-681.
- [44] Z. Luciana, B. Graça. (2008). Classification of Learning Profile Based on Categories of Student Preferences. 38th ASEE/IEEE Frontiers in Education Conference, October 22 – 25, 2008, Saratoga Springs. 978-1-4244-1970-8/08, IEEE.
- [45] Z. Luciana, R. Jose, B. Graça. (2010). An Approach to Design the Student Interaction Based on the Recommendation of e-Learning Objects. SIGDOC, September 27-29, 978-1-4503-0403-0, ACM.
- [46] Y. Liyang. (2007). From Traditional Web to Semantic Web, In: Introduction to the Semantic Web and Semantic Web Services. Taylor & Francis Group, LLC.
- [47] U. Siti, A. Rohiza, M. Shakirah. (2010). Ontology of Programming Resources for Semantic Searching of Programming Related Materials on the Web. 978-1-4244-6716-7/110, IEEE.
- [48] H. YUN, J. XU, M. J. XIONG. (2009). Development of Domain Ontology for E-learning Course, 978-1-4244-3930-0/09, IEEE.
- [49] F. Natalya, S. Michael, D. Stefan, C. Monica (2001) W. Ray, and Mark A. Musen.(2001), Creating Semantic Web Contents with Protégé-2000, IEEE INTELLIGENT SYSTEMS, 1094-7167/01, IEEE.
- [50] H. Mike, L. Fuhua, E. Larbi, and Y. Chunsheng. (2005), Constructing Knowledge Bases for E-Learning Using Protégé 2000 and Web Services, Proceedings of the 19th International Conference on Advanced Information Networking and Applications (AINA'05), 1550-445X/05, IEEE.
- [51] Khaled M. Fouad , Mostafa A. Nofal , Shehab Gamalel-Din and Nagdy M. Nagdy (2010). Adaptive E-Learning System based on Semantic Web and Fuzzy Clustering.

AUTHORS PROFILE



Khaled M. Fouad He received his Master degree of AI and expert systems. He is currently a PhD candidate in the faculty of engineering AlAzhar University in Egypt. He is working now as lecturer in Taif University in Kingdom of Saudi Arabia (KSA) and is assistant researcher in Central Laboratory of Agriculture Expert Systems (CLAES) in Egypt. His



current research interests focus on Semantic Web and Expert Systems.

Mostafa A. Nofal In 1991, he received his Ph. D. in Wireless Networks from Menoufia University, Egypt in collaboration with Dept. of Electronics and Computer Science, Southampton University, UK. From 1981 to 1986, he was a demonstrator at Faculty of Electronic Engineering, Menoufia University, Egypt. From 1986 to 1989, he was an assistant lecturer. From 1989 to 1991, he was a Ph. D. candidate at Dept. of Electronics and Computer Science, Southampton University, UK, where he was conducting research on wireless networks. In 1994 and 1999 he was a visitor professor and consultant in Southampton University, UK. From 2002, he worked as a professor of mobile networks, Menoufia University, Egypt. From 2005 until now, he is working as a professor at the Department of Computer Engineering at the College of Computers and Information Technology, Taif University, KSA. His research interests include wireless networks, data security, satellite communications and computer networks.



Hany M. Harb is professor of Computers and Systems Engineering Department - Faculty of Engineering AlAzhar university. Doctor of philosophy (Ph.D.), Computer Science, Illinois Institute of Technology (IIT) , Chicago , Illinois, USA, 1986 He is Chairman of Computers and Systems Engineering department, Chairman of Systems and Networks Unit in Al-Azhar university, and manager of WEB-Based Tansik program. He has supervision of many master's and doctoral degrees in the department of Systems and Computers Engineering.



Nagdy M. Nagdy is professor of engineering applications and computer systems, Department of Systems Engineering and Computer Engineering - Faculty of Engineering Al-Azhar University. He is working now in Al-Baha Private College of Science, Kingdom of Saudi Arabia (KSA) .He received his Ph.D in 1986. He has supervision of some master's and doctoral degrees in the department of Systems Engineering and Computer and Electrical Engineering

A Virtual Environment Using Virtual Reality and Artificial Neural Network

Abdul Rahaman Wahab Sait

Lecturer, Dept. of Computer Science
Alquwaya Community College, Shaqra University
Alquwaya, Kingdom of Saudi Arabia

Mohammad Nazim Raza

Lecturer, Dept. of Computer Science
Alquwaya College of Science, Shaqra University
Alquwaya, Kingdom of Saudi Arabia

Abstract—In this paper we describe a model, which gives a virtual environment to a group of people who uses it. The model is integrated with an Immersible Virtual Reality (IVR) design with an Artificial Neural Network (ANN) interface which runs on internet. A user who wants to participate in the virtual environment should have the hybrid IVR and ANN model with internet connection. IVR is the advanced technology used in the model to give an experience to the people to feel a virtual environment as a real one and ANN used to give a shape for the characters in the virtual environment (VE). This model actually gives an illusion to the user that as if they are in the real communication environment.

Keywords- component; Model; Virtual environment; Immersible virtual reality; Internet ; Artificial neural networks.

I. INTRODUCTION

The aim of this work is to develop a hybrid IVR and ANN model which will create a VE where a group of people chat together and at the same time they will feel as if they are in a real environment. We have used an IVR and ANN interface in this model to take the people into the VE. The model uses internet as a medium to connect the people at various regions. ANN plays a vital role here to give the presence of the image more real and make people be immersed into the environment [8][16]. Precisely the model will make a dream world to the person who uses it.

II. RELATED WORK

Manuel oliveria et al[2]. analyses a model for shared VE. They analyzed the technology uses for shared environment. They presented a model for the users to participate in a shared online game and social 3D environment. They have used virtual reality markup language to create the interface. In the environment the user will be represented as an avatar, which is an object in the virtual world. The user will be partially immersed in the environment. They have analyzed the issues regarding the VR architecture and did not give much detail about the design. We have taken ANN to synthesize the image and voice in the VE.

Dominic W.Massaro et al [3]. presents an idea on a system that synthesizes visual speech directly from the acoustic waveform. They have trained ANN to map the cepstral co-efficients of an individual's natural speech to the control parameters of an animated synthetic talking head. With this

technique, the animated talking head is generated from and aligned with the original speech of the talker.

Pengyu Hong et al [4]. gave an advanced approach on a real time speech – driven synthetic talking face provides an effective multimodal communication interface in distributed collaboration environments. They have introduced an algorithm for motion unit based facial motion tracking system. The algorithm achieves more robust results by using some high level knowledge models. They have used vector quantization, which is a classification based audio to visual conversion approach. They have trained ANN with Audio – visual training data for the system to produce the audio according the face animation. They have taken the use of the audio – visual database to obtain the visual information.

Yigiang Chen et al [5]. presents a way using the combination of clustering and machine learning methods to learn the correspondence between speech acoustic and MPEG-4 based face animation parameters. They have trained ANN to map the linear predictive co-efficients and some features of an individual's natural speech to face animation parameter. They have calculated linear predictive co-efficients and some parameters to obtain a useful vocal representation.

George votsis et al [1]. analysed the method for the recognition of user's emotional state of a PC user. They have conducted some experiment with the audio – visual databases with the ANN.

In this paper we describe a chosen VE for the user to communicate with other user and at the same time they will feel the environment as a real one by using the model. To make the environment more real we have taken ANN as a processing tool to present the user expressions in the VE.

III. PROPOSED MODEL OVERVIEW

A. Immersible Virtual Reality Design

IVR is the future technology which has the ability to make people to be immersed in the artificial environment. It acts as a vehicle for the user to move into their dream world.

As our requirements analysis proved the necessity to provide a system design for realistic shape visualization. It was obvious to use IVR techniques for state of the art 3D graphics display [2][15][16]. As IVR technology has reached a very mature level, it is very well suited for real industrial

applications which require a high level of robustness. The sense of touch can greatly enhance the users' sense of occurrence within VE which may effect in increased task performance. A successful approach has been passive haptics, which consists of providing physical artifacts for the users to physically touch whilst seeing their virtual representation. To benefit fully from the advantages of touch and keep the general purpose of VEs it is necessary to use active haptics or force feedback haptics[2].

In order to improve the level of perception our requirements analysis identified the additional need to utilize haptic force feedback. This gives intuitive and accurate interaction control to the system.

B. Artificial Neural Networks Design

An ANN is an information processing model that is similar to the way the biological nervous systems process information. It has the ability to formulate meaning from complicated or uncertain data. It exhibits the ability of adaptive learning [10]. It process information in a similar way the human brain does. It is composed of a large number of highly interconnected processing elements called neurons, working in parallel to solve a specific problem [11].

The fundamental form of ANN consists of three layers. A layer of input unit is connected to an intermediate layer of hidden units, which is connected to a layer of output unit. The behavior of the output unit depends on the activity of the hidden and input unit [9].

ANN architecture for our design is implemented with back propagation algorithm. Through back propagated networks, learning occurs during the training phase in which each input sample in the training set is applied to the input units and propagated forward[13]. The hidden layer in the middle is the heart of the process. Increase in the hidden layer gives more accurate result. The output unit will do certain matching work with the hidden layer to give the result. After a back propagated network has learned the correct classification for a set of inputs, it can be tested on a second set of inputs to see how well it classifies untrained samples. In our model, we have used this architecture to simulate the image and express their emotion of the user in the VE [14]. Users' voice, video and images are taken as input to the train the ANN [9][10][11][12]. Later stage

those data can be used for the security purpose to check the users' identity. In Fig 1 we have shown the simple IVR and ANN model.

C. Internet and Database

Internet is used as a pathway for the modern software to connect the people from all over the world together at one point of contact on the web. We require a high speed internet connection to run the model without any problem for the end user. We have to train ANN with the users' sample on the internet. As the interface runs on the internet it requires more efficiency to carry multimedia data without having any problem [13]. Another concern is the scalability of the system for a large number of users and a large amount of multimedia files. Ideally the multicast model is suited for group communication by using the same logical address.

Database is the heart of all software. As our requirement investigation we require the database with prosodic feature. The database should have the capability to contain the high quality video samples. Ideally it should have only genuine expressions of emotions. A relatively large number of facial expression recognition databases are widely employed to store and retrieve the audio and visual data [6][7]. These kinds of databases are very much useful to interact with ANN. In Fig 2 we have shown the process overview of the proposed model.

Security is one of the most concerning issues in any system [2]. The most prominent security mechanism is a user validation process during a log – on phase. In our model all users those who are going to share the environment should record their voice, video and images with the use of interface. By taking those data, the interface will check their identity in the future. So there is a less chance of security breach in this model.

IV. MODEL - WORKING METHODOLOGY

We have designed a prototype which will be used to take the people into the VE, ANN is used as a processing tool to give the virtual presence of the people by taking inputs like their image and video. Internet used as a medium to connect the IVR model. In the Fig 3, we gave a pictorial representation of functioning of the model.

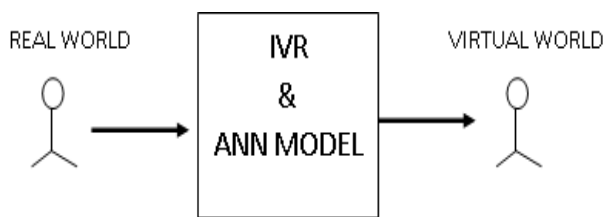


Figure 1- IVR and ANN Design

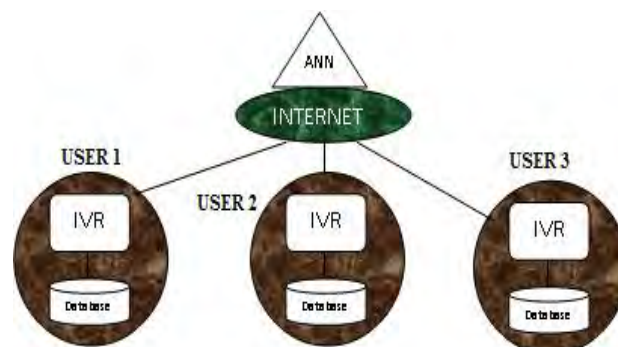


Figure 2: Process overview of the model

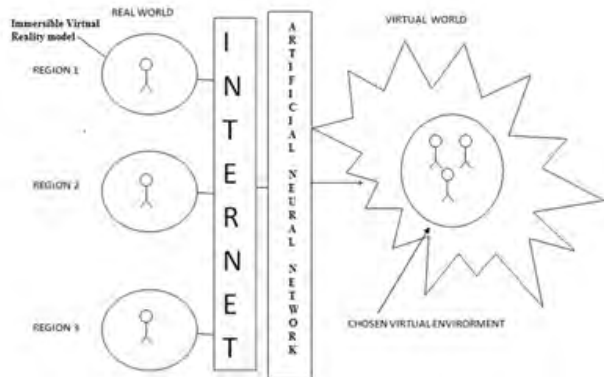


Figure 3 - Pictorial representation of functioning of the model

- Users are the participants going to share the environment. They are the manager of the model.
- Internet is the medium which connects different users together with the use of the interface. Internet speed is one of the catalysts for the model.
- ANN is the back processing tool of the model for the simulation of the user's presence in the network.

We need IVR design and ANN implemented interface and internet connection to run the model. The model has two stages, first stage is for the interface, and user has to key in details like their name, age and the region into the model. After that they have to upload their image, voice and video. These data will be stored in the audio – visual database. Later stage user data will be given to the interface to configure the user with the model. Next time when the user enters into the system

Figure 4: Illustration - 1 of the Interface

the interface will check their identity with the data already stored into it. The ANN architecture has to be trained with the samples during the training phase [13]. Here we gave illustrations of the interface in Fig 4 and Fig 5.

Figure 5: Illustration - 2 of the Interface

The second or final stage is a VE, the IVR will give the chosen environment to the users those who have participated in the network and ANN will give simulated images of the users in the VE. The interface has to be designed with some environment with the use of 3D-graphics [2][9][16]. When users select the environment the IVR design will create the environment for them. Internet is the back bone for this model. Internet speed will also important criteria for the model. ANN has the capability to recover from the failure.

V. FUTURE SCOPE AND CONCLUSION

We have presented a prototype of future technology and its methodology. We will discuss the other side of the model like internet and various design issues in the future. This hybrid VR and ANN model gives a VE in which users can meet and chat together. Further research is being conducted in order to develop a user friendly and plausible model for the user. This is the direction that we are following in our on- going research.

REFERENCES

- [1] George votsis, Nikolas D. Doulamis, Anastasios D.Doulamis, Nicolas Tsapatoulis and Stefanos D. Kollias, "A Neural – Network – Based Approach to Adaptive Human Computer Interface", ICANN 2001, LNCS 2130, PP 1054 – 1059, 2001.
- [2] Manuel oliveria, Joel Jordan, Joao Pereira, Joaquim Jorge and Anthony steed, "Analysis Domain Model for Shared Virtual Environments", The International Journal of Virtual Reality, 2009, Vol. 8, No. 4, PP 1 -30.
- [3] Dominic W.Massaro, Jonas Beskow, Michael M.Cohen, Christopher L.fry and Tony Rodriguez, "Picture my voice: Audio to Visual Speech Synthesis Using Artificial Neural Networks", Proceeding of AVSP'99, International Conference on Auditory – Visual Speech Processing, PP 133 – 138.
- [4] Pengyu Hong, Zhen Wen and Thomas S.Huang, "Real – time Speech Driven Face Animation with Expressions using Neural Network", IEEE Transactions on Neural Networks, Vol.13, Issue.4, Jul 2002, PP 916 – 927.
- [5] Yigiang Chen, Wen Gao, Zhaogi Wang and Lizao, " Speech Driven MPEG -4 Based Face Animation via Neural Network", PCM '01 Proceedings of the second IEEE pacific Rim Conference on Multimedia: Advances in multimedia information processing, PP 1108 – 1113.

- [6] Patterson E.K., Gurburz S, Tufekci Z and Gowdy J.N., "CUAVE: A New Audio – Visual Database for Multimodal Human – Computer Research", IEEE Conference on Acoustics speech and signal processing (ICASSP) 2002, PP: II 2017 – II 2020.
- [7] T.Kanade, J.F.Cohn and Y.L.Tian, " Comprehensive Database for Facial Expression Analysis", Proceeding 4th IEEE International Conference on Automatic Face & Gesture Recognition (FG '00), PP 46 – 53, 2000.
- [8] Caudell, T.P. "Application of Neural Networks to Virtual Reality.", 3rd IEEE International workshop on Robot and Human Communication, 1994, 18 – 20 Jul 94, PP 26 – 30.
- [9] John Weissmann and Ralf salomon, " Gesture Recognition for Virtual Reality Applications using Data gloves and Neural Networks", International Joint Conference on Neural Networks (1999), Vol. 1, Publisher:IEEE, PP 2043 – 2046.
- [10] M.R.Everingham, B.T. Thomas and T.Troscianko, "Head – Mounted Mobility Aid for Low Vision using Scene Classification Techniques", The International Journal of Virtual Reality (1998), Vol. 3, No.4, PP 1 – 10.
- [11] Ifat – Al – Baqee, A.S.M. Mohsin, Kurratul Ain, Mohammad Ashraf, Hossain Sadi and Md.Rakibul Hasan, "Developing a Neural Network – based Method of Faster Face Recognition by Training and Simulation", The International Journal of Engineering Science, Vol. 2, Iss. 11, PP 6694 – 6703.
- [12] Steve Lawrence, C.Lee Giles, Ah Chung Tsoi and Andrew D.Back " Face Recognition: A convolutional Neural Network Approach", IEEE Transactions on Neural Network, Special Issue on Neural Network and Pattern Recognition, PP 1 – 24.
- [13] Hagan M.T. and M.Menhaj, 1999, "Training Feed – forward Networks with the Marquardt Algorithm", IEEE Transactions on Neural Networks, Vol.5, No.6,1999, PP 989 – 993.
- [14] Lu Ye, "HRTF and Neural Network Based Prediction and Simulation Method for Indoor Sports Acoustic", International Conference on Internet Technology and Applications, 2010, 20 – 22 Aug 2010, PP 1 – 4.
- [15] Xiubo Liang, Zhen Wang, Weidong Geng and Franck Multon, " A Motion – based User Interface for the control of Virtual Humans Performing Sports", The International Journal of Virtual Reality, Vol. 10 No.3, PP 1 – 8.
- [16] Yang – Wai chow, " A Cost – Effective 3D interaction Approach for Immersive Virtual Reality", The International Journal of Recent Trends in Engineering, Vol. 1, No. 1, May 2009, PP 529 – 531.

AUTHORS PROFILE

ABDUL RAHAMAN WAHAB SAIT was born in Apr. 19, 1981 and he has completed his Masters in Information technology in 2003 in Madras University, India. Later he has done his Master of philosophy in computer science in 2007 at Periyar university, India. Now he is working as a Lecturer in Computer science, Alquwaya, Shaqra University, Kingdom of Saudi Arabia. He got interest in Virtual reality and Artificial Neural Networks. He has written many computer articles and presented paper in National conference in India.

MOHAMMAD NAZIM RAZA was born in jun. 20, 1983 and he has done his Master in Computer Application during the year 2007. Currently he is in Saudi Arabia and working as a Lecturer in Computer Science, Alquwaya, Shaqra University, Kingdom of Saudi Arabia. His interests are in Computer networks, Virtual reality and Artificial Neural networks.

Agent based Bandwidth Reservation Routing Technique in Mobile Ad Hoc Networks

Vishnu Kumar Sharma
Department of CSE, JUET, India

Dr. Sarita Singh Bhadauria
Department of Elex, MITS, India

Abstract— In mobile ad hoc networks (MANETs), inefficient resource allocation causes heavy losses to the service providers and results in inadequate user proficiency. For improving and automating the quality of service of MANETs, efficient resource allocation techniques are required. In this paper, we propose an agent based bandwidth reservation technique for MANET. The mobile agent from the source starts forwarding the data packets through the path which has minimum cost, congestion and bandwidth. The status of every node is collected which includes the bottleneck bandwidth field and the intermediate node computes the available bandwidth on the link. At the destination, after updating the new bottleneck bandwidth field, the data packet is feedback to the source. In resource reservation technique, if the available bandwidth is greater than bottleneck bandwidth, then bandwidth reservation for the flow is done. Using rate monitoring and adjustment methodologies, rate control is performed for the congested flows. By simulation results, we show that the resource allocation technique reduces the losses and improves the network performance.

Keywords- Mobile Ad hoc Networks (MANETs); Mobile Agents (MA); Total Congestion Metric (TCM); Enhanced Distributed Channel Access (EDCA); Transmission opportunity limit (TXOP).

I. INTRODUCTION

Mobile Ad Hoc networks:

The mobile ad hoc network is capable of forming a temporary network, without the need of a central administration or standard support devices available in a conventional network, thus forming a infrastructure-less network. In order to guarantee for the future, the mobile ad hoc networks establishes the networks everywhere. To avoid being an ideal candidate during rescue and emergency operations, these networks do not depend on the irrelevant hardware. These networks build, operate and maintain with the help of constituent wireless nodes. Since these nodes have only a limited transmission range, it depends on its neighboring nodes to forward packets [1].

II. RESOURCE ALLOCATION AND ITS ISSUES IN MANET

For the sake of improving and automating the quality of service of the networks, efficient resource allocation techniques are required. Resource allocation is carried out in a static manner on the hours to months scale of time in telecommunication networks. If traffic varies significantly, then resource allocated in the statical manner is inadequate or under-exploited. [2]

Wireless networks are emerging hastily and endlessly with the condition related to the rising transmission speeds, number

of users and services. Owing to the huge number of the customers, the resource on the network has high requirement and competition. In addition, the necessity of the resources in a wireless networks varies depending on the network load and radio channel conditions. The resource reservation inefficiency causes high losses to the service providers and results in inadequate user proficiency. Hence active resource management system which has capability to make best use of resource is required. [3]

For the ad hoc network application, the bandwidth reservation process are required for the real time flows. If the admission control is matched with the network characteristics, then reservations can avoid congestion occurrence. Best effort traffic is not restricted to any scheme, and thus can intersect on the bandwidth share of the advantaged traffic resulting in assurance more delicate. The feasible solution to this difficulty will be assigning a constant bandwidth for best effort traffic. The solution doesn't take resources needed for the traffic or topology of the network into consideration. These solution frequently results in optimal use of the network resources. The another option is bandwidth allocation to best effort traffic depending on the topology and bandwidth existing in every mobile. [4]

The active nature of the MANET causes unpredicted intrusion of attacks or faults which further results in seperation of the network, performance degradation, violation of the QoS requirements and more specifically disturb the bandwidth reservation. [10]

III. PREVIOUS WORKS

In paper [11], we proposed an agent based congestion control technique in MANET. In our technique, the node is classified in one of the four categories depending on whether the traffic belongs to background, best effort, video or voice AC respectively. Then MA estimates the total congestion metric by calculating the queue length and the channel contention and it is applied to the routing protocol to select the minimum congested route.

In paper [12] we proposed an agent based power control technique in MANET. In power control technique, the nodes are chosen based on the power level. The nodes with maximum power level are selected as listening nodes (LN) which will always be in active node and remaining nodes are selected as non-listening nodes (NLN) which will awake in periodic manner. The status of LN nodes keeps changing in every time cycle. The source transmits the data packets to the destination through the selected path. If the node receiving the packet is awake, the packet is transmitted to that node otherwise node

checks for the nearest listening for transmitting the packet. In this manner, the packets are transmitted in hop-by-hop manner with reduced power consumption.

In our existing approaches, though mobile agents reduce the congestion and power, inefficient allocation of the resources may incur heavy losses to the service providers as well as poor user experience. Hence in our extension work, we are planning to include the agent based resource allocation in MANET.

IV. RELATED WORKS

R.Gunasekaran et al [5] proposed the high-privileged and low-privileged architecture (HPLP) for Ad Hoc network for achieving optimal differentiated services for different classes of users. The new protocol, D-MACA, was implemented. Among the various factors influencing the differentiated services, bandwidth reservation is only considered and different factors that can influence the efficiency of the bandwidth reservation are identified. The drawback of this proposed approach is that the complexity issues such as processing time, transaction time (latency), buffer management and memory utilization is not considered.

Kumar Manoj et al [6] proposed an algorithm that contains bandwidth calculation and slot reservation for mobile networks which could be applied to multimedia ad hoc wireless networks. Specially, the bandwidth information can be used to assist in performing the handoff of a mobile host between two base stations. Traffic flows with different QoS types have been considered. In addition to, standby routing enhances the performance in the mobile environment.

Rafael Guimarães et al [7] proposed a QoS reservation mechanism for multirate ad hoc wireless networks that allows bandwidth allocation on a per flow basis. By multirate they refer to those networks where wireless nodes are able to dynamically switch among several link rates. This allows nodes to select the highest possible transmission rate for exchanging data, independently for each neighbor. This reservation approach provides a feasible way to avoid congestion, guaranteeing, thus QoS requirements to ongoing connections.

Maria Canales et al [8] proposed an adaptive cross-layer architecture based on the cooperation between a QoS routing protocol and the MAC level. This joint operation allows to perform a distributed admission control capable of providing the required end-to-end QoS adapting the operation to the characteristic variant environment of MANETs. The proposed scheme has been designed a flexible parameters configuration that allows to adapt the system response to the observed grade of the mobility in the environment.

Wang Xiangli et al [9] proposed a distributed bandwidth reservation protocol (DBRP) for QoS routing in ad hoc networks. The protocol adopts a TDMA-based model, derives from AODV, refers to the idea of three slot states, and adopts two-time reservation and controlling-flooding scheme. The protocol takes both the hidden-terminal and exposed-terminal problems into account. And it can solve the simultaneous reservation of several paths. In addition, controlling-flooding

method can effectively control routing overhead, and two-time reservation can improve request success rate.

V. PROPOSED WORK

A. Overview

The mobile agent from source starts forwarding the data packets through the path containing minimum cost, congestion and bandwidth availability. The packets upon reaching every intermediate node updates its list with the node information such as its id, flag, power level, node activating counter, information about the neighbor node, cumulative assigned rates for incoming and outgoing flow and requested data rate stored in the bottleneck bandwidth (BW_{BN}) field and the intermediate node computes the available bandwidth (B_{av}) on the link. If $B_{av} > BW_{BN}$, then the node forwards the packet to the next node on the path. Else the node replaces BW_{BN} with B_{av} and proceeds to forward to the next node. When the data packet reaches the destination, BW_{BN} field is copied to new packets and feedback to the source. The intermediate node updates its routing table with new BW_{BN} value when the data packet is traversing towards the source. The source after receiving the data packet updates its routing table with the new BW_{BN} value. If the $B_{av} > BW_{BN}$, then reservation of bandwidth for the flow can be proceeded. Otherwise, the BW_{BN} is overwritten with the B_{av} . The rate control technique concentrates on rate monitoring and adjustment methodologies where the cumulative assigned rate for incoming and outgoing flow helps in rate adjustment.

B. Available Bandwidth estimation

Every node is in charge for estimating the available bandwidth on its link. For a given node,

Let B_{av} = available bandwidth.

L = link capacity associated with one-hop neighbor i .

ACA be the cumulative assigned rates for all incoming and outgoing flows.

Hence the sum of the assigned incoming and outgoing flow rates and available bandwidth on the link should be equal the capacity of the link i . This can be expressed as

$$ACA^i + B_{av_i} = L_i$$

The mobile agent from the source node forwards the data packet along a given path towards the destination. The data packet constitutes the requested bandwidth value stored in the bottleneck bandwidth field. Each intermediate node is responsible for determining whether or not sufficient bandwidth is available on the local outgoing link to support the new flow request.

The link capacity is measured and available bandwidth is defined by

$$B_{av_j} \triangleq \max \{0, L_j - ACA^j\}$$

C. Resource allocation technique

1) 1. Entries of node's routing table

Each node constitutes the routing table that includes the entries of its id, flag, power level, node activating counter, cumulative assigned rates for incoming and outgoing flow and requested data rate stored in the bottleneck bandwidth (BW_{BN}) field. The amount of quantity of the routing table entries is found

based on the number on the active incoming and outgoing flows which is expressed as $n(n-1)$, where n is the number of neighbors of the node.

The routing table also includes the following values

Assigned ACA^{ij} corresponding to incoming and outgoing flow.

Counter CNT_{ij} for the number of bits that have arrived in the current measurement window.

Measured rate CA^{ij} from the previous measurement window.

Every node is responsible for policing the incoming and outgoing flow to the cumulative assigned rate ACA^{ij} . This measured rate CA^{ij} helps in performing rate-adjustment.

The source node selects path with minimum power consumption and congestion as per previous paper [12]. The following section describes the steps involved in the bandwidth reservation technique.

2) 2. Steps Involved in Bandwidth Reservation Step 1

The mobile agent from the source node forwards the data packet that contains the IP address of source and destination, flow ID and requested data rate stored in the BW_{BN} field to the destination.

Step 2

The intermediate node upon receiving the data packets determines the B_{av} on its outgoing link.

Step 3

If B_{av} is greater than the BW_{BN} value, then

Node forwards the packet to the next node on the path

Else

Node replaces the BW_{BN} field with the value of B_{av} and forwards the packet to next node.

End if

This process continues till the data packet reaches the destination.

Step 4

When the destination node receives the data packet, it copies the value of the BW_{BN} to the new data packet and sent back to the source node using the reverse path.

Step 5

The intermediate node upon receiving the data packet updates its routing table with the new BW_{BN} and then forwarded the packet to the next node.

Step 6

The routing table is updated in the following way. Let n_i (represented as in-hop) be the next node to which the new data packet will be sent and n_j (represent as out-hop) be the node from which the packet was received.

If the routing table entry for incoming and outgoing flow already exists (i.e the flow is active).

Then current BW_{BN} value in the new data packet is added to the reserved rate CA_{ij} , associated with the incoming and outgoing flow.

Else

The routing table entry is created with an assigned rate value CA_{ij} (set equal to the BW_{BN} value of the feedback data packet).

End if

This process continues till the data packet reaches the source.

Step 7

When the data packet reaches the source node, the source establishes the real-time flow based on the value of the BW_{BN} field.

If the value of B_{av} in source node is greater than or equal to the BW_{BN} value in the packet

Then reservation of bandwidth for the flow can proceed

Else

The BW_{BN} value in the new data packet is overwritten with the (smaller) value B_{av} .

end if

D. Rate Monitoring and Adjustment

In the rate monitoring strategy for a real time flow, the rate of flow is measured and compared with the assigned rate which is updated in the routing table. If the rate measured is lesser than the reserved rate by the sufficient margins, then the reserved rate is reduced by certain factor.

The traffic rate of a given flow during time interval t can be measured by rate monitoring methodology. This is achieved by maintaining a counter that keeps the count value of the total number of bits arriving on an incoming and outgoing flow over a time t . As each packet arrives on a given flow (i, j) , a counter CNT_{ij} is incremented in terms of the size of the packets (in

bits). After lapse of time period t , the measured rate CA^{ij} becomes

$$CA^{ij} = CNT_{ij}/t$$

The following step describes the rate adjustment strategy.

If $(ACA^{ij} - CA^{ij}) > x$, then

$$ACA^{ij} = ACA^{ij} - (1 - \gamma^x)$$

end if

If $ACA^{ij} < Th$, then

The flow is removed from the routing table.

end if

Here $\gamma \in (0, 1)$ represents a design parameter. x represents certain percentage.

V. SIMULATION RESULTS

A. Simulation Model and Parameters

We use NS2 [13] to simulate our proposed technique. In the simulation, the channel capacity of mobile hosts set to the same value: 11Mbps. In the simulation, mobile nodes move in a 1000 meter x 1000 meter region for 50 seconds simulation time. Initial locations and movements of the nodes are obtained using the random waypoint (RWP) model of NS2. It is assumed that each node moves independently with the same average speed. All nodes have the same transmission range of 250 meters. The node speed is 5 m/s. and pause time is 5 seconds.

TABLE 2. THE SIMULATION SETTINGS AND PARAMETERS

No. of Nodes	50
Area Size	1000 X 1000
Mac	802.11e
Radio Range	250m
Simulation Time	50 sec
Routing Protocol	AODV
Traffic Source	CBR and Video
Video Trace	JurassikH263-256k
Packet Size	512
Mobility Model	Random Way Point
Speed	5m/s
Pause time	5 sec
MSDU	2132
Rate	50kb,100kb,.....250Kb
No. of Flows	4,5,6,7 and 8
Initial Energy	5.1 J
Transmit Power	0.360 w
Receiving Power	0.395 w
Idle Power	0.335 w

B. Performance Metrics

We compare the performance our Agent based Bandwidth Reservation (ABR) technique with the BRAWN [7] scheme. The performance is evaluated mainly, according to the following metrics.

- **Aggregated Bandwidth:** We measure the received bandwidth for class1 (VBR) and class2 (CBR) traffic of all flows

- **Fairness Index:** For each flow, we measure the fairness index as the ratio of throughput of each flow and total no. of flows
- **Total Bandwidth:** It is the sum of received bandwidth of class1 and class2.

VI. RESULTS

A. Effect of Varying Rates

In the initial experiment, we measure the performance of the proposed technique by varying the rate as 50,100, 150, 200and 250Kb.

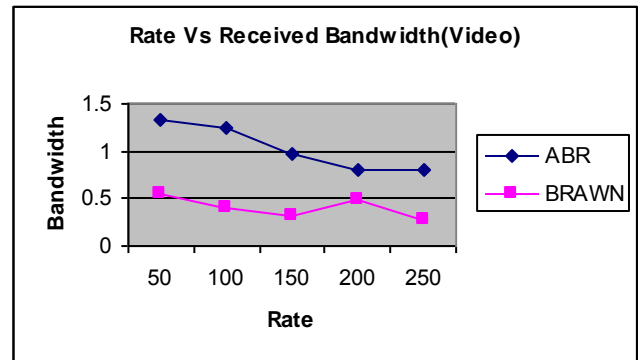


Fig 1: Rate Vs Received Bandwidth

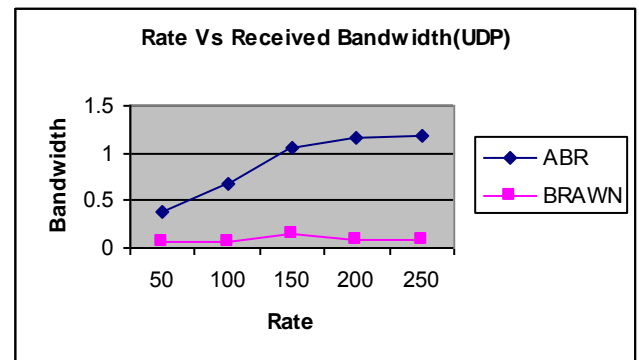


Fig 2: Rate Vs Received Bandwidth

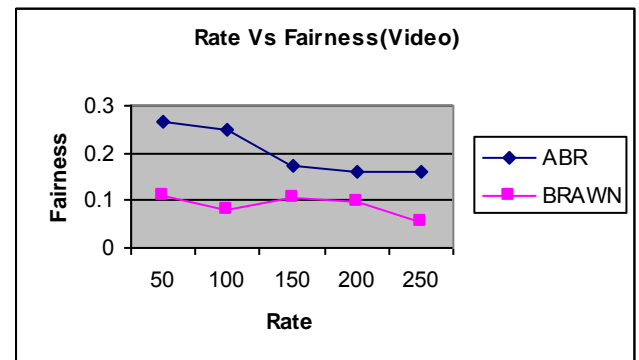


Fig 3: Rate Vs Fairness

Fig: 1 and Fig: 2 give the aggregated bandwidth for Video and UDP traffic. From the figures, it can be seen that ABR has received more bandwidth when compared with BRAWN.

Fig: 3 and Fig: 4 give the fairness index for Video and UDP traffic. From the figures, it can be seen that ABR achieves more fairness when compared with BRAWN.

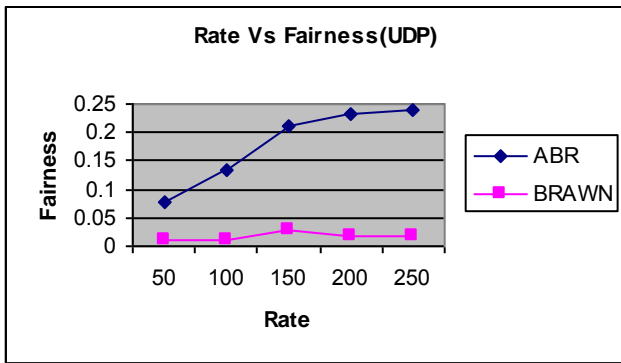


Fig 4: Rate Vs Fairness

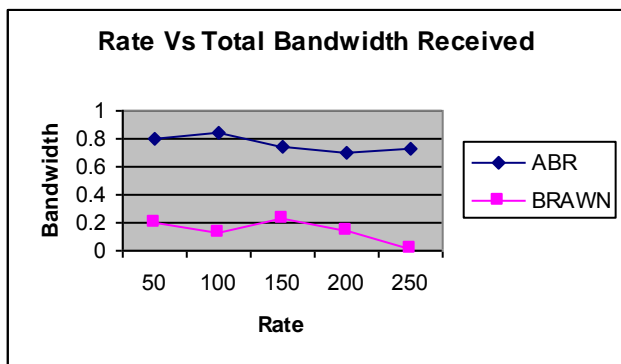


Fig 5: Rate Vs Total Bandwidth

Fig: 5 give the Total Bandwidth ratio. From figure, we can see that the proposed ABR has high total Bandwidth ratio than the BRAWN

B. Effect of Varying Flows

In the next experiment, we compare our proposed technique by varying the number of flows as 4,5,6,7 and 8.

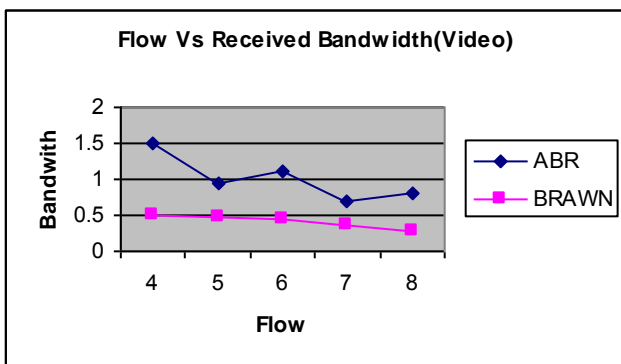


Fig 6: Flow Vs Received Bandwidth for VBR flows

Fig: 6 and Fig: 7 give the aggregated bandwidth for Video and UDP traffic. From the figures, it can be seen that ABR has received more bandwidth when compared with BRAWN.

Fig: 8 and Fig: 9 give the fairness index for Video and UDP traffic. From the figures, it can be seen that ABR achieves more fairness when compared with BRAWN.

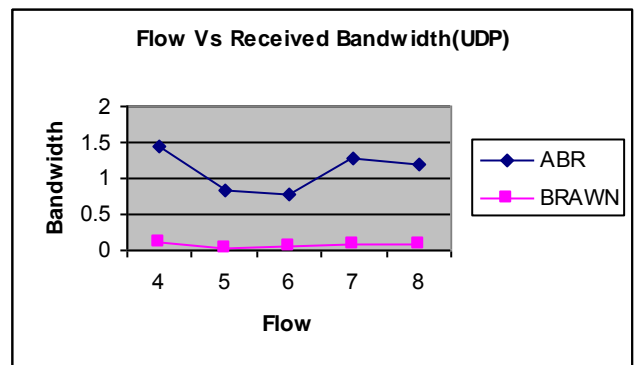


Fig 7: Flow Vs Received Bandwidth for CBR flows

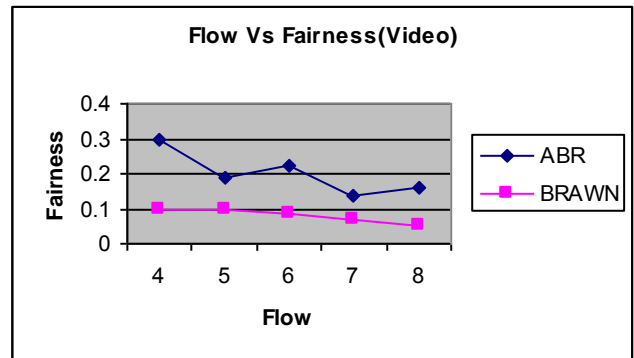


Fig 8: Flow Vs Fairness for VBR flows

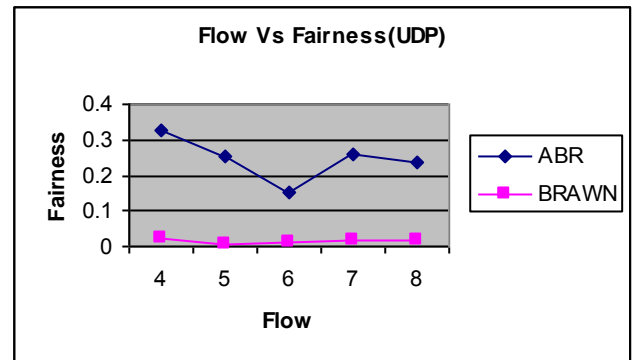


Fig 9: Flow Vs Fairness

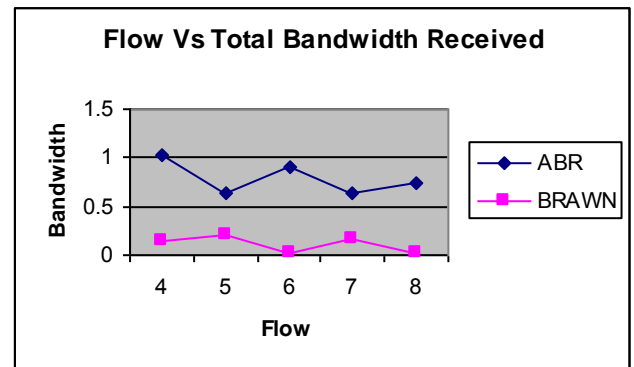


Fig 10: Flow Vs Total Bandwidth

Fig: 10 give the Total Bandwidth ratio. From figure, we can see that the proposed ABR has high total Bandwidth ratio than the BRAWN.

VII. CONCLUSION

In this paper, we have proposed an agent based bandwidth reservation technique for MANETs. The mobile agent from the source starts forwarding the data packets through the path which has minimum cost, congestion and bandwidth. The status of every node is collected which includes bottleneck bandwidth (BW_{BN}) field and the intermediate node computes the available bandwidth (B_{av}) on the link and finally the packet is intended to destination. After updating the new BW_{BN} field, the data packet is feedback to the source. In resource reservation technique, if the B_{av} is greater than BW_{BN} , then bandwidth reservation for the flow proceeds. Otherwise the BW_{BN} field is overwritten with B_{av} value. The rate control technique is added that contains traffic policing and rate monitoring and adjustment. The cumulative assigned rate of the incoming and outgoing flows helps in rate adjustments. By simulation results, we have shown that the resource allocation technique reduces the losses and improves the network performance.

REFERENCES

- [1] S.Santhosh baboo and B.Narasimhan, "A Hop-by-Hop Congestion-Aware Routing Protocol for Heterogeneous Mobile Ad-hoc Networks", International Journal of Computer Science and Information Security, 2009
- [2] A. Capone, J. Elias, F. Martignon, G. Pujolle, "Dynamic resource allocation in communication networks", in: Proceedings of Networking 2006, Coimbra, Portugal, 15-19 May 2006, also published in Springer LNCS, vol. 3976, 2006, pp. 892-903.
- [3] Babbar, R., Fapojuwo, A., Far, B., "Agent-Based Resource Management in Hybrid Wireless Networks", IEEE Canadian Conference on Electrical and Computer Engineering, Vol. 3 (2004), pp. 1297-1300.
- [4] C. Chaudet, I. Guérin Lassous and J. Zerovnik, "A distributed algorithms for bandwidth allocation in stable ad hoc networks", IFIP International Conference on Wireless On-Demand Network Systems (WONS), pp. 101-115, January 2004.
- [5] R. Gunasekaran and V. Rhymend Uthariaraj, "Differentiated Bandwidth Allocation in Mobile Ad Hoc Networks (MANET) – A Profile Based Approach", Asian Journal of Information Technology, 2007.
- [6] Kumar Manoj Member, IAENG, S. C. Sharma & S.P. Singh, "Dynamic Behavior of Bandwidth Control Management in Mobile Ad-Hoc Network", Proceedings of the World Congress on Engineering and Computer Science, Vol I, 2009.
- [7] Rafael Paoliello-Guimarães, Llorenç Cerdà, José M. Barceló, Jorge García-Vidal, Michael Voorhaen, Chris Blondia, "Quality of service through bandwidth reservation on multirate ad hoc wireless networks Ad Hoc Networks", 388-400, 2009.

- [8] Maria Canales, Jose Ramon Gallego, Angela Hernandez Solana, Antonio Valdovinos, "QoS provision in mobile ad hoc networks with an adaptive cross-layer architecture", Journal Wireless Networks, Volume 15 Issue 8, November 2009.

- [9] Wang Xiangli, Li Layuan, Gong Berican, Wang Wenbo, "A Distributed Bandwidth Reservation Protocol for QoS Routing in Mobile Ad Hoc Networks", international conference on wireless communications, networking and mobile computing, vol 1-15, 2007.



- [10] Binod Kumar Pattanayak, Manoj Kumar Mishra, Alok Kumar Jagadev, and Manoj Ranjan Nayak, "A Cluster-based QoS Support To Bandwidth Preservation With Concept Of Survivability In Multi-hop Mobile Ad Hoc Networks", Proceedings of The World Congress on Engineering and Computer Science, pp247-254, 2009.

- [11] Bhadauria SS, Sharma V., Framework and Implimentation of an Agent Based Congestion Control Technique for Mobile Ad-hoc Network, ICAC3 2011, CCIS, Volume 125, Springer, pp. 318-327, 2011.

- [12] Sharma V, Bhadauria SS, "Agent Based Congestion Control Routing for Mobile Ad-hoc Network" Wimon-2011, CCIS, Springer, Volume 197, pp.324-333, 2011.



- [13] Network Simulator, <http://www.isi.edu/nsnam/ns>

- [14] Vishnu Kumar Sharma and Dr. Sarita Singh Bhadauria. Congestion and Power Control Technique Based on Mobile Agent and Effect of Varying Rates in MANET. IJACSIT Vol 1 No 1.

AUTHORS PROFILE

Vishnu Sharma, working as a faculty in CSE Dept. Of Jaypee University of Engg and Technology and presently, he is a Ph.D Candidate at M.P. Technical University, He has about 13 years of teaching experience. He has published more than fifteen papers in the area of Mobile Ad-hoc Networks & Mobile Computing at National/International Level. He has published the books on Mobile Computing and Advanced Mobile Computing in UPTU, Lucknow (UP) and IP University, New Delhi. He is a life time member of International Association of Computer Science and Information Technology (IACSIT), Singapore and member of CSI. His area of interest includes Mobile Computing, Computer Networks and Communication Technologies, Cryptography and Network Security.

Dr. Sarita Singh Bhadauria She is a Professor in Department of Electronics MITS, Gwalior (M.P.) India; she has about 25 years of teaching and research experience. She has published more than fifty papers in the area of Digital Image Processing, Computer Networks, Mobile Ad-hoc Networks, Mobile Communication and Cryptography and Network Security and Digital Communication at National/International Level. She is a life time member of ISTE, IETE, IEEE, CSI and HAM RADIO. Her areas of interest include Wireless Communication and Digital Image Processing, Computer Networks, and Mobile Ad-hoc Networks, Mobile Communication and Cryptography and Network Security and Digital Communication and Communication Technologies.

Sensor Node Deployment Strategy for Maintaining Wireless Sensor Network Communication Connectivity

Shigeaki TANABE, Kei SAWAI, Tsuyoshi SUZUKI

Tokyo Denki University,
School of Engineering, Department of Information and Communication Engineering,
Tokyo, Japan

Abstract—We propose a rescue robot sensor network system in which a teleoperated rescue robot sets up a wireless sensor network (WSN) to gather disaster information in post-disaster underground spaces. In this system, the rescue robot carries wireless sensor nodes (SNs) and deploys them between gateways in an underground space on demand by the operator's command to establish a safe approach path before rescue workers enter. However, a single communication path only is setup, because the rescue robot linearly deploys SNs between gateways. Hence, the rescue robot cannot be operated remotely if the communication path is disconnected by, for example, SN failure or changes in the environmental conditions. Therefore, SNs must be adaptively deployed so as to maintain WSN communication connectivity and negate such situations. This paper describes an SN deployment strategy for construction of a WSN robust to communication disconnection, caused by SN failure or deterioration of communications quality, in order to maintain communication connectivity between SNs. We thus propose an SN deployment strategy that uses redundant communication connection and ensures communication conditions between end-to-end communications of the WSN. The proposed strategy maintained communication conditions such that throughput between end-to-end communications in the WSN. Experimental results verifying the efficacy of the proposed method are also described.

Keywords-wireless sensor network; deployment strategy; communication connectivity

I. INTRODUCTION

In recent years, disaster mitigation measures have been discussed in order to reduce the damage caused when a disaster occurs [1]. For disaster mitigation, various action plans have been devised and implemented in attempts to reduce damage and to facilitate early post-disaster rehabilitation and reconstruction. In particular, in an actual disaster zones, fire crews and rescue teams still must actually enter a post-disaster site to determine the current extent of the damage, because flexible responses are required according to the damage conditions. However, activities in such post-disaster situations present a high risk of personal injury to rescue workers due to secondary disasters, such as fire, or the collapse of walls and ceilings. Therefore, to reduce such risks, information gathering for early detection of possible threats is important in preventing secondary disasters, and actions to mitigate further disaster can be planned based on this information. Thus, continuous and exhaustive gathering and monitoring of information is necessary in the disaster zone in order to detect a secondary disaster rapidly, because secondary disasters can be caused by a number of factors that vary over time and locations.

Disaster information-gathering systems, for example, in the form of artificial satellites, unmanned aerial vehicles, and balloon flights are mainly used to gather disaster information over a wide area [2]. However, employment of these systems in

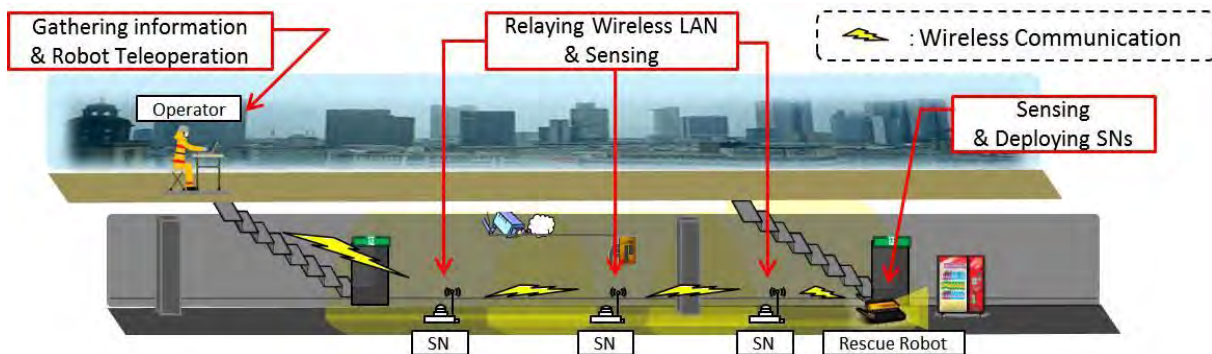


Figure 1. Conceptual sketch of information-gathering system in an underground space using RRSN

underground spaces is difficult, and viable alternative methods for gathering information in such spaces have yet to be proposed. Moreover, many cases have been reported in which problems have actually been caused by existing disaster-prevention equipment that has failed to operate during a disaster. Therefore, there is an increased need for information-gathering systems to determine the post-disaster status for disaster mitigation in underground spaces, such as subway stations and underground malls in urban areas. Underground spaces are particularly significant since, owing to their relative structural integrity, they are often expected to be used as evacuation shelters and as stores for emergency supplies.

Based on the above considerations, we have proposed a rescue robot sensor network (RRSN) system. In this system, a teleoperated mobile robot, such as a high-mobility rescue robot, sets up a wireless sensor network (WSN) to gather disaster information in post-disaster underground spaces (Fig. 1). WSNs consist of a number of small devices called sensor nodes (SNs), each of which is equipped with wireless communication functionality, various sensors, a processor, and a power source. A WSN is then a network system that can communicate and use data mutually gathered by the spatially distributed SNs. An ad hoc network that connects each SN one-by-one can be constructed by deploying a large number of SNs, and such a network is very easily enhanced compared with wired and fixed networks. WSNs can thus provide various services by collecting and processing the information acquired by the SNs, and WSNs technology is expected to be applicable in many fields, such as in cooperative monitoring of environmental conditions over large areas, in communication in sites where the construction of such infrastructure is difficult, and in information gathering for disaster-relief support.

Although establishing a wireless connection is generally difficult between ground level and an underground space, RRSN is able to create a communication link from a base station at ground level to a rescue mobile robot in an underground space via a wireless ad hoc network that connects each SN. Previously, we developed RRSN prototypes and verified their functionality experimentally [3]-[6]. In the proposed RRSN scenario [3], only a single communication path is setup because the rescue mobile robot linearly deploys SNs between the gateways in an underground space in order to establish a safe approach path for rescue workers. Hence, the rescue mobile robot cannot be remotely operated if a communication route is disconnected as a result of battery drain, SN failure, or changes in environmental factors, such as obstacles, radio interference, or radio wave conditions. Therefore, SNs must be adaptively deployed to maintain WSN communication connection in order to handle such unforeseeable circumstances. In addition, preserving the functionality of an information-gathering network is important, and maintenance is achieved by continually determining the state of the WSN and recovering the communication path if necessary. However, complex tasks for WSN restoration are not possible during the time-sensitive situation at a disaster site. Hence, to continue smooth teleoperation of rescue mobile robot under actual circumstances, an SN deployment method robust to disconnection of communication is vital. This paper thus proposes an SN deployment method for construction of a WSN

robust to communication disconnection caused by SN failure or deterioration of communications quality. Experiments testing SN deployment and the maintenance of communications quality when using the proposed method are also described. These experiments verify the validity of the proposed.

The rest of the paper is organized as follows. Section 2 outlines works related to this study. Section 3 describes our system configuration and explains the proposed method of WSN construction. Section 4 evaluates the proposed method and presents the experimental results. Finally, concluding remarks are given in Section 5.

II. RELATED WORKS

Recently, many SN deployment strategies have been discussed in the WSN research field. In these strategies, deployment methods have been proposed based on evaluation scales that consider factors such as packet routing, energy efficiency, power-saving, and coverage area.

Chen et al. proposed an improved WSN routing protocol designed to maintain communication quality between SNs by referring only to the received signal strength indication (RSSI) value [7]. Iranli et al. have studied energy efficient strategies for deployment to construct a two-level WSN [8]. In that study, a control method for energy consumption in 2-hop networks is developed by utilizing an existing routing protocol. Zhang and Hou proposed a method for maintaining sensor coverage and communication connectivity by utilizing a WSN with a minimal number of SNs [9]. Furthermore, Zhang suggests a method of optimal geographical density control in large scale sensor networks that minimizes the number of SNs.

In the research into development of a mobile SN for construction of the WSN, Dantu et al. developed the Robomote, which added mobility to the Mica Mote. The Robomote can install WSN algorithms easily using the Mica Mote [10]. The MICAbot is a similar system to the Robomote, but it has higher mobility [11]. Suzuki et al. discussed WSN protocol-based research into a method where the mobile SN carries the data by physical movement when the WSN is disconnected [12][13].

Several SN deployment methods using mobile SNs and mobile robots to construct the WSN have been developed [14]-[16]. Parker et al. proposed a WSN construction method using an autonomous helicopter for environmental monitoring and urban search and rescue [17]. Umeki et al. proposed an ad-hoc network system, Sky Mesh, using a flying balloon for targeted disaster rescue support [18]. Also, deployment methods have been developed based on virtual interaction between the SNs based on several physical models, such as the potential field model and the fluid flow model [19][20].

However, none of these methods considered end-to-end network connectivity or network performance measures, such as throughput. Many researches are premised on communication link being maintained automatically. Moreover, evaluations in these methods are based solely on computer simulations. Hence, implementation of a practical wireless connection system was not attempted, and applying the methods is nontrivial for typical rescue mobile robot utilizing 2.4 GHz-band wireless communications.

III. SN DEPLOYMENT STRATEGY

A. SN deployment conditions

To develop a test system for evaluating the proposed strategy, the conditions for SN deployment by RSSN are defined as follows.

Firstly, RSSN gathers environmental information in the vicinity of the gateways connecting the ground level and a first basement level, because this information is necessary for rescue workers to enter into the underground space. Under the Japanese Building Standards Law, safety stairs leading directly to ground level must be installed a maximum of every 30 min an underground space for escape during a disaster. As a result, our SNs are linearly deployed over a distance of about 40 m, including the stairs and the passage between the gateways. Environmental information gathered by the RSSN is then continually transmitted to the operator, who can ascertain the disaster status within the underground space from this information.

The SN-loaded robot enters into the basement from the ground-side gateway through teleoperation by an operator at the base station. As the robot progresses it measures the communication conditions between the base station and its current position. When communication conditions satisfy the requirements specified by a deployment strategy, the robot indicates this information to the operator who decides an SN deployment point by referring to these conditions. The operator then gives a deployment command to the robot at the relevant location and the robot places an SN on the floor. The robot thus continues to measure the communication conditions between deployed SNs while moving toward a destination gateway, placing each SN in the same manner to setup the WSN.

Here, we assume that the extent of structural collapse is limited and that the environmental structure is largely maintained, because the underground space has earthquake resistance due to its structural integrity [21]. Therefore, the robot is not impeded in the space by unexpected obstacles resulting from the disaster. The robot and SNs are equipped with the wireless local area network (LAN) Institute of Electrical and Electronics Engineers (IEEE) standard, 802.11b. This standard is tolerant to communication disruption from obstacles and can connect to SNs within 50 m in 1-hop communication.

B. Required specifications of communication

The robot cannot be operated remotely via the WSN when the wireless communication quality has deteriorated between SNs. The wireless communication quality of the WSN constructed between the base station and the moving robot must always be maintained for smooth teleoperation of the robot. To satisfy this condition, a throughput of about 1.0 Mbps is required for robot teleoperation to allow the passing of WSN data traffic [22]. However, for end-to-end multi-hop communication in a wireless ad hoc network, the throughput decreases as the number of SNs increases. For cases having greater than five hops, a throughput of 1.0 Mbps is not ensured due to the upper limit of the IEEE 802.11b standard being exceeded [23]. Thus, consideration of the number of SNs used in WSN communication is important for maintaining the necessary throughput. In the proposed RSSN, a WSN with up to 4-hop communication is constructed by the base station, SNs, and the robot (Fig. 2). The communication distance connected by these SNs (including the base station and robot) satisfies the above condition in terms of the IEEE 802.11b specifications.

However, the throughput of end-to-end communication changes depending on the throughput between each SN. The distance between SNs cannot exceed a certain limit, since the throughput between two SN decreases as their distance increases. In contrast, when the distance between SNs is shortened, the WSN cannot cover the intended environment, because the number of SNs is restricted. In particular, the above communication condition is not satisfied when the SNs are deployed based on the nominal communication distance in wireless communication specifications, because the wireless communication performance dynamically changes depending on the environment in a post-disaster underground space. Therefore, determining the appropriate interval between SNs is important in order to cover our assumed environment while maintaining the required throughput. From the abovementioned requirements, we propose an SN deployment strategy that includes the following considerations:

- Determination of an appropriate SN interval.
- Setup of a WSN with a robust and fault-tolerant communication.

The following section describes an algorithm for determination of SN deployment positions that takes account of

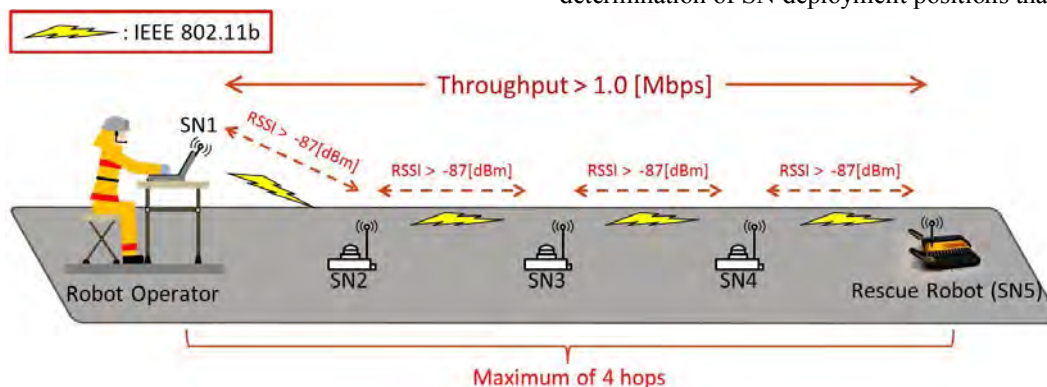


Figure 2. System configuration for SN deployment model that considers communication conditions

the above condition of deploy SNs while maintaining a throughput of 1.0 Mbps.

C. SN deployment model based on communication conditions

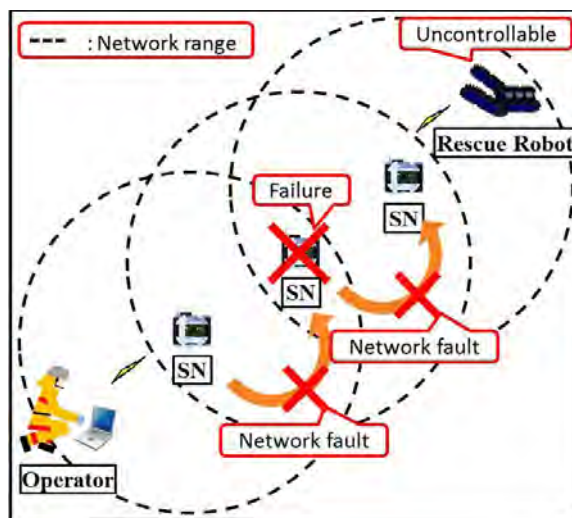
To ensure communication performance between SNs, the robot measures throughput while transporting the SNs to their required locations. Thus, the robot confirms that the SNs are able to maintain a 1.0 Mbps throughput for end-to-end communication, and deploys SNs while ensuring the throughput between them. For 4-hop WSN communication, a 7.0 Mbps throughput must be maintained between each SN to guarantee a 1.0 Mbps throughput for end-to-end communication. However, an auto-fallback system has been installed in the IEEE 802.11b standard that adjusts throughput automatically in response to changes of RSSI value. This system adjusts the upper limit of the throughput to less than 7.0 Mbps when the RSSI value falls below -86 dBm. Therefore, both the throughput and RSSI values must be sustained to ensure the required communication performance, and the robot measures both the throughput and RSSI value between each SN accordingly. The robot moves continuously to the destination gateway while maintaining the 1.0 Mbps of throughput required for end-to-end communication and an RSSI value of -86 dBm between each SN. When the RSSI value starts to decay, the robot places the SN before the RSSI values fall below -86 dBm (Fig. 2). The robot thus sets up the WSN by repeatedly placing SNs on the floor, one by one, by measuring the throughput and RSSI value between SNs. This method is expected to provide SN deployment and smooth robot teleoperation adaptive to changes in communication performance caused by environmental interference.

D. SN deployment model based on redundant communication connection

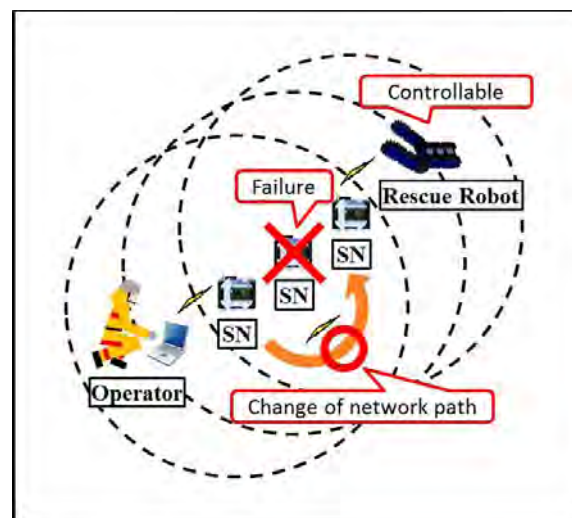
Because the disconnection of communication caused by SN failure is a severe problem in setting up a WSN, several research studies have attempted to improve the fault tolerant performance of WSNs [24][25]. In conventional methods, a large number of SNs are deployed for WSN construction, and the communication path then changes according to SN failure. In our system, however, these methods cannot be used owing to our restricted number of SNs and linear deployment design. Thus, we instead propose a redundant placement method for SNs. That is, two SNs are placed within an area where the RSSI value from last SN is greater than -86 dBm and throughput is greater than 7.0 Mbps. Upon failure, automatic restoration of the communication path to the robot is then possible by connecting to the next running SN, and the 1.0 Mbps of throughput between the end-to-end communications in the WSN is maintained if the network routing is changed (Fig. 3).

E. SN deployment strategy based on both models

We lastly propose an SN deployment strategy that uses redundant communication connection and ensures communication conditions between end-to-end communications of the WSN. In this section, an SN deployment algorithm using the robot based on above-mentioned models is described.



(a) SN deployment without considering redundant communication connection



(b) SN deployment with considering redundant communication connection

Figure 3. SN deployment model considering redundant communication connection

Figure 4 shows the steps of the algorithm.

- i) As shown by path (1) in Fig. 4, the SN-loaded robot moves from base station to a destination gateway by teleoperation. While moving, the robot measures both the RSSI value between the base station and current position, and the throughput between end-to-end communications.
- ii) When the throughput or RSSI value starts to degrade, the robot stops before they reach their relevant thresholds (1.0 Mbps and -86 dBm, respectively). The robot measures distance it has moved from the start point (P2 in Fig.4) and stores this stop position.
- iii) The robot transmits the communication conditions data to the operator. The operator confirms the information and sends an SN deployment command.
- iv) The robot moves back half of the distance it has traveled (path (2)), and places first SN (SN2) at point P1.

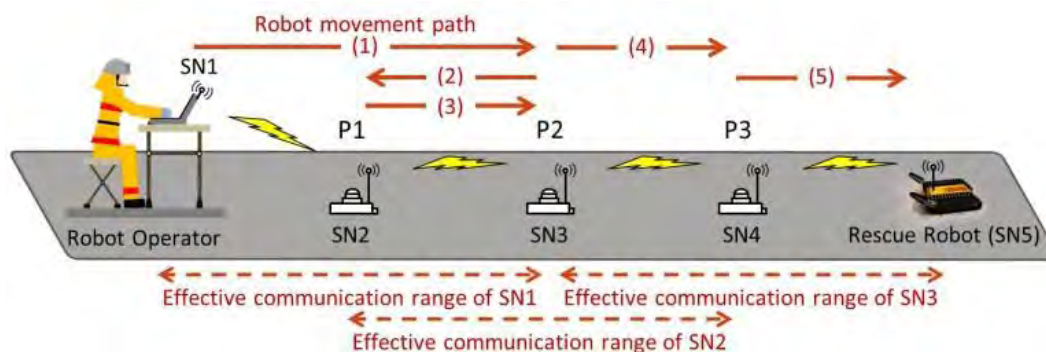


Figure 4. Steps in WSN deployment algorithm

- v) After that, the robot moves forward to the stored stopping point P2 and places SN3, and again measures the throughput between the base station and current position to confirm the end-to-end communication conditions.
- vi) In a similar manner, the robot continues to move forward (path (4)), measuring the communication conditions between SN2 and its current position. The robot stops before the communication performance degrades below the thresholds and places SN4 at P3.
- vii) The robot continues moving in this way (path (5)) until it has arrived at the destination gateway.

IV. EVALUATION OF SN DEPLOYMENT STRATEGY

A. Experiment for confirming SN communication conditions

We evaluated our proposed development strategy in a passageway of 40 m in length. In this evaluation, we measured the throughput to determine the communication quality between the base SN and a destination SN.

The SN deployment interval was decided by RSSI; that is, the electrical field density between SNs. The threshold level for maintaining communication quality was defined as -86 dBm, and the extended distance of the WSN was calculated based on sum of the SN intervals. The RSSI was measured every 5 m that the robot moved.

Table I shows the SN specifications. Armadillo-300 (At-Mark Techno Ltd.) was used as the SN controller. For RSSI measurements, we used the “wlanconfig” command contained in the Wireless Tools package in the Debian Linux distribution. To measure the throughput of a packet, “utest” (NTT Communications Ltd.) was used.

The crawler-type mobile robot (TOPY industry Ltd.) in Fig. 5 was adopted for use as the rescue mobile robot in this experiment.

24 bytes data containing operation modes and each actuator velocity were transmitted as the robot operation commands. The communication status measurement commands described above were also transmitted with parameters. 67 bytes data containing each actuator velocity, current values, a tilt angle of the robot and a range sensor data were received as the robot status. The RSSI value and the throughput were also received as the current communication conditions.

Image data was not used in this evaluation, but when the throughput of 1.0 Mbps is maintained between end-to-end communications, it can be communicated enough.

Figure 6 shows the experimental results. The maximum extended distance of the WSN was 95 m under the conditions of 4-hop communication, and RSSI and throughput values exceeding -86 dBm and 1.0 Mbps, respectively. Therefore, we established that the proposed strategy could be applied in our assumed environment of 40 m.

TABLE I. SN SPECIFICATIONS

Devices and tools	
Linux Controller	Armadillo-300
Wireless LAN	IEEE 802.11b (2.4 GHz)
Measurement Software	Wireless tools and utest

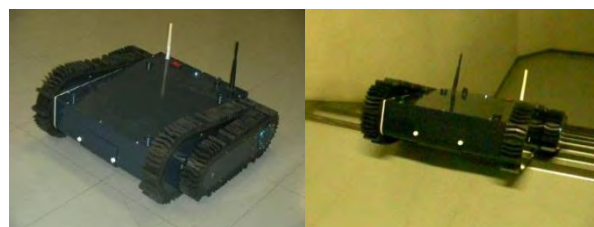


Figure 5. Crawler-type mobile robot

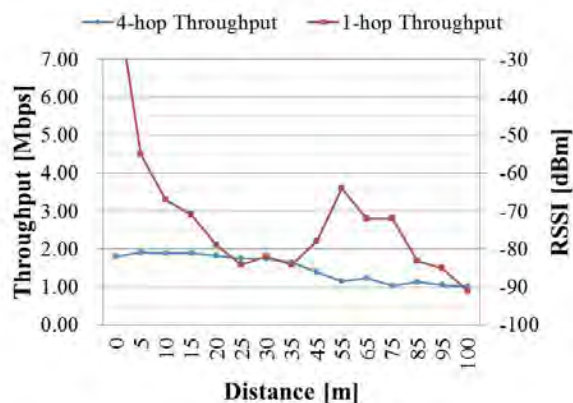


Figure 6. Example measurement results for throughput and RSSI

B. Evaluation of SN deployment strategy

We measured the WSN communication conditions in order to evaluate the fault-tolerant capability of the proposed strategy. For this evaluation, we compared two deployment strategies. Strategy 1 determined SN placement based on the maximum range described in the communication specifications. Strategy 2 was the proposed strategy, where SN positions were determined by referring the maximum extended distance shown in Fig. 6. For WSN construction using the proposed strategy, by referring Fig. 6, the SN placement positions were calculated as 0, 35, 65, 85, and 95 m. The power supply of one SN was then randomly switched off to simulate SN failure. The experimental environment was same as in the previous experiment.

When SN failure occurred, the communication path changed in both strategies, and the throughput and RSSI values between the SNs in new communication path were measured. Tables II and III shows the results when SN2, SN3, and SN4 fail while using Strategies 1 and 2, respectively. For Strategy 1 the communication conditions decayed to -87 dBm or less and 1.0 Mbps or less. However, for Strategy 2, communication conditions were maintained above -87 dBm and 1.0 Mbps. Therefore, both communication maintenance and failure resistance in using the proposed strategy are verified.

Although the proposed strategy maintains communication connectivity against the communication disconnection caused by SN failure, measurement results of the communication conditions change depending on the environment condition.

The communication link of the RRSN is disconnected if the communication conditions are degraded dramatically. For such case, the other strategy must be applied to solve communication disconnection. For example, we have proposed WSN reconstruction method that the robots deploy additional or alternate SNs to necessary sites to restore communication links [3]. When a wireless communication cannot be available, the strategy to gather as much information as possible using wired mobile robot is also considered.

TABLE II. EXPERIMENTAL RESULT USING THE MAXIMUM EXTENDED WSN BASED ON COMMUNICATION DISTANCE SPECIFICATIONS (STRATEGY 1)

Fault SN	RSSI [dBm]	End-to-End Throughput [Mbps]
SN2	-89	Immeasurable level
SN3	-87	0.267
SN4	-91	0.507

TABLE III. EXPERIMENTAL RESULTS USING PROPOSED DEPLOYMENT STRATEGY (STRATEGY 2)

Fault SN	RSSI [dBm]	End-to-End Throughput [Mbps]
SN2	-79	1.132
SN3	-86	1.041
SN4	-81	1.377

V. DISCUSSION AND FUTURE DIRECTION

This paper has focused on a SN deployment strategy considering the communication conditions and redundancy. In this paper, the key result is that SN deployment positions can be determined adaptively according to the communication conditions, and communication connectivity can be maintained against the communication disconnection caused by SN failure. However, it is difficult to apply this prototype system to actual post-disaster situations. Therefore, we are now developing new crawler robots for SN deployment and information gathering in post-disaster situations based on the findings obtained from our studies. The proposed strategy in this paper will be applied to new robots. These component technologies are integrated, and the performance of the RRSN must be further developed and improved to enable it to be applied practically in assumed environment to gather information on the effects of disasters in underground spaces.

VI. CONCLUSION

This paper has proposed a WSN deployment strategy that maintains communication conditions and has a fault-tolerant communication connection. The proposed strategy maintained communication conditions such that throughput between end-to-end communications in the WSN enables smooth teleoperation of a mobile robot in a post-disaster underground space. Experimental results showed the effectiveness of the proposed strategy.

This strategy, which prevents communication disconnection caused by SN failure, is considered effective for WSN deployment in actual disaster scenarios, because the rapid implementation of actions to mitigate secondary disasters in disaster zones requires the stable referral of disaster information. We will apply the proposed strategy to WSN deployment in practical underground spaces in the future.

ACKNOWLEDGMENT

This work was partially supported by the Research Institute for Science and Technology of Tokyo Denki University, Grant Number Za10-01 / Japan.

REFERENCES

- [1] Y. Kawata, "The great Hanshin-Awaji earthquake disaster: damage, social response, and recovery," Journal of Natural Disaster Science, Vol. 17, No. 2, pp.1-12, 1995.
- [2] H. Kawakata, Y. Kawata, H. Hayashi, T. Tanaka, K. C. Topping, K. Yamori, P. Yoshitomi, G. Urakawa and T. Kugai, "Building an integrated database management system of information on disaster hazard, risk, and recovery process" Annuals of Disas. Prev. Res. Inst., Kyoto Univ., No.47 C, 2004.
- [3] T. Suzuki, R. Sugizaki, K. Kawabata Y. Hada and Y. Tobe , "Autonomous deployment and restoration of sensor network using mobile robots," International Journal of Advanced Robotic Systems, ISSN 1729-8806, Vol. 7, No. 2, June 2010, pp.105-114
- [4] H. Sato, K. Kawabata, T. Yugo, H. Kaetsu and T. Suzuki, "Wireless camera nodes deployment by a teleoperated mobile robot for construction of sensor network," ICROS-SICE International Joint Conference 2009 (ICCAS-SICE2009), pp.3726-3730, August 18-21, Fukuoka International Congress Center, Fukuoka, JAPAN, 2009

- [5] K.Sawai, T.Suzuki, H.Kono, Y.Hada and K.Kawabata, "Development of a sensor node with impact-resistance capability for gathering disaster area information," 2008 International Symposium on Nonlinear Theory and its Applications (NOLTA2008), pp.17-20 (A1L-C2), ISBN: 978-4-88552-234-5, Budapest, Hungary, September 7-10, 2008.
- [6] H. Sato, K. Kawabata and T. Suzuki, "Information Gathering by wireless camera node with Passive Pendulum Mechanism," *International Conference on Control, Automation and Systems 2008 (ICCAS2008)*, pp.137-140, October 14-17, Seoul, Korea, 2008.
- [7] F. Chen, C. Wu, P. Ji and Y. Zhang, "A communication quality improved routing protocol for wireless sensor network", *Automation and Logistics, 2009. ICAL '09. IEEE International Conference on Digital Object Identifier*, pp. 616 – 620, 2009.
- [8] A. Iranli, M. Maleki, M. Pedram, "Energy efficient strategies for deployment of a two-level wireless sensor network," *Proceedings of the 2005 International Symposium on Low Power Electronics and Design 2005 (ISLPED '05)*, pp. 233-238, Southern California Univ., Los Angeles, CA, USA, August 8-10, 2005.
- [9] H. Zhang, J. Hou, "Maintaining sensing coverage and connectivity in large sensor networks", *Ad Hoc and Sensor Wireless Networks*, Vol. 1, No. 1-2, 2005.
- [10] K. Dantu, M. Rahimi, H. Shah, S. Babel, A. Dhariwal, and S. S. Gaurav., "Robomote: enabling mobility in sensor networks," *Fourth International Symposium on Information Processing in Sensor Networks (IPSN2005)*, pp. 404-409, 2005.
- [11] M. B. McMickell, B. Goodwine and L. A. Montestruque, "MICAbot: A robotic platform for large-scale distributed robotics," *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA2003)*, 2, 1600-1605, 2003.
- [12] R. Suzuki, K. Makimura, H. Saito and Y. Tobe, "Prototype of a sensor network with moving nodes," *Special Issue on INSS 2004, Transaction SICE, E-S-1*, pp. 52-57, 2006.
- [13] R. Suzuki, K. Sezaki and Y. Tobe, "A Protocol for policy-based session control in disruption tolerant sensor networks. Special section on Ubiquitous Sensor Networks, IEICE TRANSACTIONS on Communications, E90-B(12), pp. 3426-3433, 2007.
- [14] M. A. Batalin, and G. Sukhatme, "Sensor coverage using mobile robots and stationary nodes," *Proceedings of the SPIE (SPIE2002)*, No. 4868, pp. 269-276, 2002.
- [15] S. Miyama, M. Imai and Y. Anzai, "Rescue robot under disaster situation: position acquisition with omni-directional sensor," *Proceedings of 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2003)*, pp. 3132-3137, 2003.
- [16] M. Sugano, T. Kawazoe, Y. Ohta and M. Murata, "An indoor localization system using RSSI measurement of a wireless sensor network based on the ZigBee standard," *Proceedings of the sixth IASTED International Multi-Conference on Wireless and Optical Communication*, pp. 504-508, 2006.
- [17] L. E. Parker, B. Kannan, F. Xiaoquan and T. Yifan, "Heterogeneous mobile sensor net deployment using robot herding and line of sight formations," *Proceedings of 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2003)*, Vol. 3, pp. 2488-2493, 2003.
- [18] T. Umeki, H. Okada and K. Mase, "Evaluation of Wireless channel quality for an ad hoc network in the sky, SKYMESH," *The Sixth International Symposium on Wireless Communication Systems 2009 (ISWCS'09)*, pp.585-589, 2009.
- [19] A. Howard, M. J. Matric and G. S. Sukhatme, "Mobile sensor network deployment using potential fields: A distributed, scalable solution to the area coverage problem," *Distributed Autonomous Robotics Systems 5*, Springer-Verlag, pp. 299-308, 2002.
- [20] M. R. Pac, A. M. Erkmen and I. Erkmen, "Scalable self-deployment of mobile sensor networks: A fluid dynamics approach," *Proceedings of 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2006)*, pp. 1446-1451, 2006.
- [21] P. J. Godard, "Urban Underground Space and Benefits of Going Underground", *Proceedings of World Tunnel Congress 2004 and 30th ITA General Assembly*, pp.1-9, Singapore, pp. 22-27, May 2004.
- [22] J. Yamashita, K. Sawai, Y. Kimitsuka, T. Suzuki, Y. Tobe, "The design of direct deployment method of sensor nodes by utilizing a rescue robots in disaster areas," *SICE Annual Conference 2008*, pp.183, 2B3-4, 6 Dec. 2008 (in Japanese)
- [23] K. SAWAI, T. SUZUKI, "Evaluation of network construction method considering reduce of throughput by utilizing rescue robot," *Proceedings of the 2010 JSME Conference on Robotics and Mechatronics*, No. 10-4, 1A2-C09(2), 2010 (in Japanese)
- [24] M. Ishizuka, M. Aida, "Stochastic Node Placement Improving Fault Tolerance in Wireless Sensor Networks," *Institute of Electronics, Information, and Communication Engineers, J88-B(11)*, pp. 2181-2191, 2005.
- [25] B. Krishnamachari, S. Iyengar, "Distributed Bayesian algorithms for fault-tolerant event region detection in wireless sensor networks," *IEEE TRANSACTIONS ON COMPUTERS*, Vol. 53, No. 3, pp. 241-249, March 2004.

AUTHORS' PROFILE

Shigeaki TANABE - Graduate school student at Tokyo Denki University. His research interests include sensor network communication and robot-assisted sensor networks.

Kei SAWAI - Assistant professor at Tokyo Denki University. His research interests include networked robotics and robot-assisted sensor networks.

Tsuyoshi SUZUKI - Professor at Tokyo Denki University. His research interests include multi-robot systems, human-robot interaction, telerobotic systems, networked robotics, and robot-assisted sensor networks.

Detection and Extraction of Videos using Decision Trees

Sk.Abdul Nabi

Dept. of CSE
AVN Inst. of Engg. & Tech.
Hyderabad, India

Shaik Rasool

Dept. of CSE
S.C.E.T.
Hyderabad, India

Dr.P. Premchand

Dept. of CSE
University College of Engineering,
OU
Hyderabad, India

Abstract— This paper addresses a new multimedia data mining framework for the extraction of events in videos by using decision tree logic. The aim of our DEVDT (Detection and Extraction of Videos using Decision Trees) system is for improving the indexing and retrieval of multimedia information. The extracted events can be used to index the videos. In this system we have considered C4.5 Decision tree algorithm [3] which is used for managing both continuous and discrete attributes. In this process, firstly we have adopted an advanced video event detection method to produce event boundaries and some important visual features. This rich multi-modal feature set is filtered by a pre-processing step to clean the noise as well as to reduce the irrelevant data. This will improve the performance of both Precision and Recall. After producing the cleaned data, it will be mined and classified by using a decision tree model. The learning and classification steps of this Decision tree are simple and fast. The Decision Tree has good accuracy. Subsequently, by using our system we will reach maximum Precision and Recall i.e. we will extract pure video events effectively and proficiently.

Keywords- DEVDT; Data Processing; Data Pre-Processing; Decision Tree and Training Data.

I. INTRODUCTION

Over a period of time, data researchers have shown immense interest in the study of data mining. This is quite natural, as the database field started in the commercial community and this community still has much influence over the types of questions being studied. Digital multimedia differs from previous forms of combined media in that the bits that represent text, images, animations, audio, video and other signals can be treated as data by computer programs. One fact of this diverse data in terms of underlying models and formats is that it is synchronized and integrated. Hence, it can be treated as integral data records. Virtual communities (in the broad sense of this word, which includes any communities mediated by digital technologies) are another example where generated data constitutes an integral data record. Such data may include data about member profiles, the content generated by the virtual community and communication data in different formats including email, chat records, SMS messages and video conferencing records. Not all multimedia data is so diverse. An example of less diverse but larger in terms of the collected amount is the data generated by video surveillance systems [2 , 4] where each integral data record roughly

consists of a set of time-stamped images – the video frames. In any case, the collection of such integral data records constitutes a multimedia data set. The challenge of extracting meaningful patterns from such data sets has led to the research and development in the area of multimedia data mining.

Multimedia databases are widespread and multimedia data sets are extremely large. There are tools for managing and searching within such collections but the need for tools to extract hidden useful knowledge embedded within multimedia data is becoming critical for many decision-making applications. The tools needed today are tools for discovering relationships between data items or segments within images, classifying images based on their content, extracting patterns from sound, categorizing speech and music, recognizing and tracking objects in video streams, relations between different multimedia components and cross-media object relations [1]. The overall design of a multimedia database differs markedly from that of a standard textual database. Browsing and querying in the former environment utilizes entities and attributes that are usually hidden from the casual user. Since many of the data mining tasks in standard databases concern associations between different attributes in a multimedia environment, the nature of the attributes over which associations are constructed becomes quite important.

II. MULTIMEDIA DATA MINING MODELS

Multimedia data mining is a challenging field due to the non-structured nature of multimedia data [5]. Such ubiquitous data is required, if not essential in many applications. Multimedia database design differs distinctly from that of a standard textual database. Browsing and querying in the former environment utilizes entities and attributes that are usually hidden from the casual user. Since many of the data mining tasks in standard databases concern associations between different attributes. In a multimedia environment, the nature of the attributes over which associations are constructed becomes quite important. For example, it is possible to mine a rule of the form like this: consumers who remove Brand A paper towels from the display and examine the package for at least 20 seconds will also purchase it [s%, c%] [6] using textual information which only appears in relational tables. This would entail having purchasing information entered into a back-end database via a point-of-sale terminal as well as a person observing for how long people examine various

products and entering this information in the same database. However, this rule can also be mined by content-based retrievals of video(s) taken of the shopping experience of various shoppers during some period of time. For this latter approach, it is not obvious what attributes of the overall database are being used for the mining task. In order to construct a more detailed characterization of the different sorts of data mining in a multimedia database environment, we now address the general notion of multimedia data models.

There are many multimedia data models in the literature [7, 8]. However, all such data models are similar at a high enough level of abstraction. They all should represent the following types of information

1. The detailed structure of the various multimedia objects.
2. Structure dependent operations on multimedia objects.
3. Multimedia objects properties.
4. Relationships between multimedia objects and real-world objects.
5. Portions of multimedia objects that have representation relationships with real-world objects, the representation relationships themselves and the methods used to determine them.
6. Properties, relationships and operations on real-world objects.

Modeling the structure of a multimedia object is important for many reasons, not the least of which is that various operations are defined on these objects which depend on its structure. These operations are used to create derived multimedia objects for similarity matching (edge maps) as well as various composite multimedia objects from individual component multimedia objects (multimedia presentations). An example of a multimedia object property is the name of the object. For example, 'Titanic' is the name of a particular video object. A relationship between a multimedia object and a real-world object would be the stars-in relationship between the actor Victoria Foyt and the video Titanic. Suppose that Titanic ship is a real-world object being represented in the database and that a particular region of frame six of the video Titanic is known to show this object. This small portion of the byte span of the entire video is also considered to be a first-class database object, called a semcon [9, 10], for iconic data with semantics. Both the many-one relationship represents which holds between this semcon and Titanic ship object, as well as the many-many relationship appearing-in which holds between the Titanic ship object and the video Titanic should be captured either implicitly or explicitly by any multimedia data model.

These relationships enable metadata mediated browsing. Such behavior is exhibited when one clicks a mouse whose cursor is over a semcon representing a Titanic ship. As a database object, this man-made structure is represented in the database by a tuple in the Monument table. Doing a join, we may then get tuples representing information concerning the designers of this structure. Finally, we may view images of containing pictures of these people. The relationship represents enables the database system to navigate from the

semcon of the Titanic ship to the database tuple representing this structure, while the relationship appearing-in enables the system to navigate from tuples concerning the designers of the Titanic ship to their images. Semcons, as first-class database objects and have attributes. These attributes include various features extracted from them that can be used for similarity matching over other multimedia objects. Features should be first-class database objects as well and include such things as color histograms and texture maps. In a multimedia database environment, querying consists of utilizing semcons for searching for multimedia objects corresponding to the same real-world object. If two semcons have similar features, the semcons themselves are similar.

III. DECISION TREE MODEL

A Decision tree is a flowchart-like tree structure where each internal node (non-leaf node) denotes a test on an attribute, each branch represents an outcome of the test and each leaf node (terminal node) holds a class label. The topmost node in a tree is the root node. The example of decision tree is shown in figure:

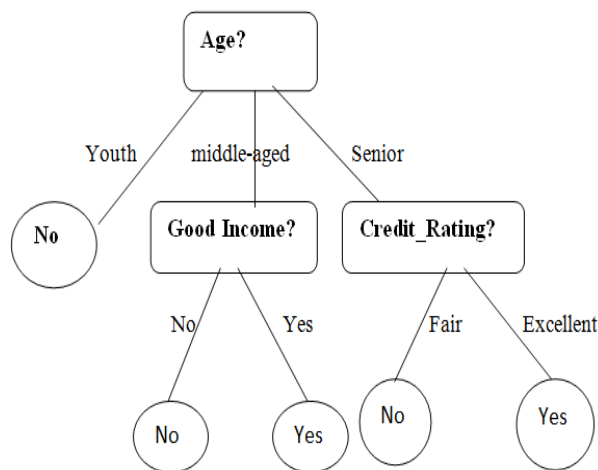


Figure: 1 Decision Tree for purchase_house

An example (shown in the figure: 1) represents the decision tree for purchasing the House (Own house), in which each internal node represents a test on an attribute. Each leaf node represents a class (i.e. either purchase_house = yes or purchase_house = no).

A decision tree [11] is a decision-making device which assigns a probability to each of the possible choices based on the context of the decision: $P(f/h)$, where f is an element of the future attributes (the set of choices) and h is a history (the context of the decision). This probability $P(f/h)$ is determined by asking a sequence of questions $q_1 q_2 \dots q_n$ about the context, where the i^{th} question asked is uniquely determined by the answers to the $i - 1$ previous questions. Each question asked by the decision tree is represented by a tree node and the possible answers to this question are associated with branches emanating from the node. Each node defines a probability distribution on the space of possible decisions. A node at which the decision tree stops asking questions is a leaf node. The leaf nodes represent the unique states in the decision-making problem, i.e. all contexts which

lead to the same leaf node have the same probability distribution for the decision.

The power of decision-tree model is not in their expressiveness but instead in how they can be automatically acquired for very large modeling problems. The decision-tree learning algorithm increases the size of a model only as the training data allows. The decision-tree learning algorithm increases the size of a model only as the training data allows. The leaf distributions in decision trees are empirical estimates, i.e. relative-frequency counts from the training data. Unfortunately, they assign probability zero to events which can possibly occur. Therefore, it is necessary to smooth empirical decision-tree models. For that we have considered C4.5 Decision Tree System, which is used for classification from a set of Trained Data.

C4.5 builds decision trees from a set of training data in the same way as ID3 using the concept of information entropy. The training data is a set $S = s_1, s_2, \dots$ of already classified samples. Each sample $s_i = x_1, x_2, \dots$ is a vector where x_1, x_2, \dots represent attributes or features of the sample. The training data is augmented with a vector $C = c_1, c_2, \dots$ where c_1, c_2, \dots represent the class to which each sample belongs. At each node of the tree, it chooses one attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. Its criterion is the normalized information gain (difference in entropy) that results from choosing an attribute for splitting the data. The attribute with the highest normalized information gain is chosen to make the decision. This algorithm then recurs on the smaller sub lists.

IV. PROPOSED SYSTEM

In this paper, we have proposed a new system DEVDT for detection and extraction of video events by using Decision Tree classifiers (shown in fig.2). The aim of this system is to detect the relevant and pure events and extracted portion of video events effectively and efficiently.

The training data for data mining is the multimodal features (visual and audio) extracted for each video event. It is event-based because video events are the basic indexing unit for video content analysis [12, 13, 14]. In addition, we have adopted an advanced video event detection method, having the advantage of producing some important visual features and mid-level features (e.g., object information) during event detections. However, due to the small percentage (e.g., 1%) of the positive samples with the huge amount of negative samples domain knowledge utilizing visual and audio clues has been used in our data pre-filtering step to clean the original feature data set in order to provide a reasonable input training data set for the data mining component. To our best knowledge, there is hardly any work addressing this issue. Finally, the decision tree model generated by the data mining process will be tested and the overall performance is evaluated by using large amounts of long video sequences with different styles and produced by different broadcasters. By using our DEVDT system, we will reach up to 92% for both Recall and Precision.

The architecture of our system is shown in Figure 2. As can be seen from this figure, the proposed framework consists of the following three major components: Video Processing, Data Pre-Processing and Data Mining.

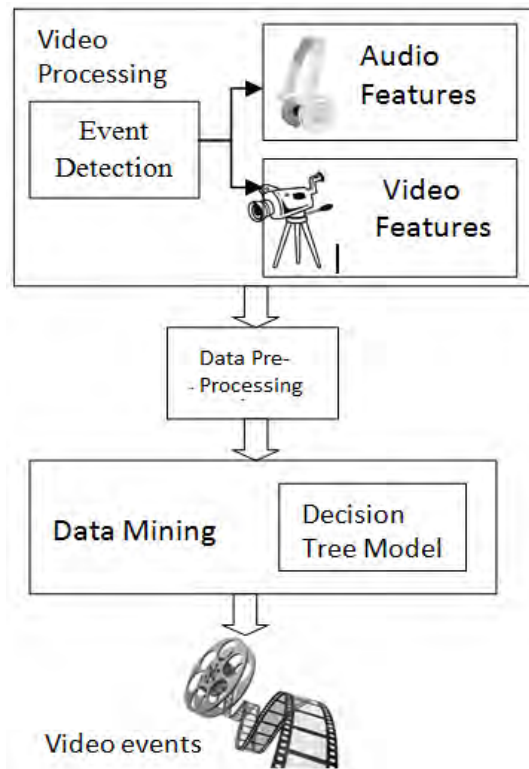


Figure: 2 DEVDT (Detection & Extraction of Video Events using Decision Tree)

A. Video Processing

Parse the raw video sequences by using a video event detection subcomponent. It not only detects video event boundaries but also produces some important visual features during event detection. The detected event boundaries are passed to feature extraction, where the complete multimodal features (visual and audio) are extracted for each event.

B. Data Pre Processing

Use domain knowledge such as visual audio clues to eliminate the noise data and reduce the irrelevant data from the original feature set since the ratio of actual events over the non-related events is very small (e.g., 1 actual event out of 100 events). By data pre- Processing, the ratio of positive samples over negative samples can be increased to 1:20.

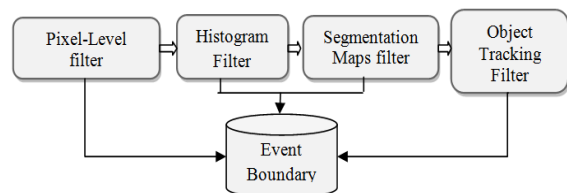


Figure 3: Data Pre-Processing for video event detection

C. Video Event Detection

The first step for video processing and the detected event boundaries is the basic unit for video feature extraction. In this

System, we have proposed a Data Preprocessing (multi-filtering architecture) including the pixel-level comparison, histogram comparison and segmentation map techniques (as shown in Figure 3). The first two filters can compensate for each other in reducing the numbers of both false positives and false negatives. In addition, since the object segmentation and tracking techniques are much less sensitive to luminance change and object motion. They are used as the last filter in this multi-filtering architecture to help determine the actual event boundaries. The advantages of this method are: Firstly, It has high precision and recall values. This overall performance is considered based on more than 1,500 testing events approximately and then secondly it can generate a set of important visual features for each event during the process of event detection. Thus the computation for extracting visual features can be greatly reduced.

D. Visual Event Extraction

In addition to event boundaries, the process of video event detection also generates a rich set of visual features associated with each video event. Among these visual features, pixel-change represents the average percent of the changed pixels between frames within an event which is output by the first filter (Pixel-Level Filter). The feature histogram change indicates the mean value of the histogram difference between frames within an event and is output by the second filter (Histogram Filter). Both of the two global features are important indications for camera motions and object motions. Other mid-level features such as the mean (back-mean) and the variance (back-var) values of the background pixels can be obtained via the segmentation filter.

E. Audio Feature Processing

Both time-domain and frequency-domain audio features are considered in our framework. Since the semantic meaning of an audio track is better represented by the audio features of a relatively longer period. We also explore both the clip-level and shot-level audio features. In this study, we define an audio clip with a fixed length of one second, which usually contains a continuous sequence of audio frames.

The generic audio features are divided into three groups: volume features (volume), energy features (energy), and Spectrum Flux features (sf). For each generic audio feature, the audio files are processed to obtain the audio features at both clip-level and shot level. The audio data is sampled at a sampling rate of 16,000 HZ. An audio frame contains 512 samples, which lasts 32ms under a sampling rate of 16,000 HZ. Within each clip, the neighboring frames overlap 128 samples with each other. In order to model the energy properties more accurately, four energy sub-bands are also used in this study. In this process we have used all IO audio features (i.e.1 volume feature, 5 energy features, and 4 spectrum flux features) to improve the performance.

F. Mining Video Events using Decision Trees

In our DEVDT system, the decision tree logic is adopted for mining events in videos. In this phase, we will take the 'cleaned' feature data as the training data and build a decision tree model suitable for video event detection. An interior node in a decision tree involves testing a particular attribute and the

branches that fork from that node correspond to all possible outcomes of a test. Eventually, a leaf node is formed which carries a class label that indicates the majority class within the final partition. The classification phase works like traversing a path in the tree. Starting from the root, the instance's value of a certain attribute decides which branch to go at each internal node. Whenever a leaf node is reached its associated class label is assigned to the instance. The algorithm exploited in this study is adopted from the C4.5 decision tree [3].

In the decision tree generation process, the information gain ratio [15] criterion is used to determine the most appropriate attribute for partitioning due to its efficiency and simplicity. Numeric attributes are accommodated by a two-way split, which means one single breakpoint is located and serves as a threshold to separate the instances into two groups. The voting of the best breakpoint is based on the information gain value.

C4.5 is an algorithm used to generate a decision tree developed by Ross Quinlan. It is an extension of Quinlan's earlier ID3 algorithm. The decision trees generated by C4.5 can be used for classification. In pseudo code, the general algorithm for building decision trees is [16]:

- 1) Check for base cases
- 2) For each attribute a
- 3) Find the normalized information gain from splitting on a
- 4) Let a_{best} be the attribute with the highest normalized information gain
- 5) Create a decision node that splits on a_{best}
- 6) Recurse on the sub lists obtained by splitting on a_{best} , and add those nodes as children of node

The advantages of using Decision Tree are it doesn't require any domain knowledge or parameter setting. Therefore it is appropriate for exploratory knowledge. C4.5 Decision trees can handle high dimensional data. It can be used for both continuous and discrete attributes. In order to handle continuous attributes it creates a threshold and then splits the list into those whose attribute value is above the threshold and those that are less than or equal to it. The learning and classification steps of this Decision Tree are simple and fast. Accurate results can be obtained with this Decision Tree method.

V. CONCLUSION AND FUTURE ENHANCEMENTS

This paper reviewed the importance of multimedia data mining and concludes that one of the major issues of multimedia data mining is the accuracy and efficiency of getting results for extraction of audio and video events from raw data. In this paper we have discussed about the multimedia mining modals and the decision tree logic concepts. We have proposed a framework DEVDT, Which uses data mining concept of Decision Tree classifier model for detection and extraction of video and audio events to improve efficiency and accuracy of extraction.

The construction of Decision Tree is performed by recursively partitioning the training set with respect to certain criteria until all the instances in a partition have the same class label or no more attributes can be used for further partitioning.

This can be done by using C4.5 Decision Tree system. The advantage of using this Decision tree system is to handle both continuous and discrete attributes. It can also handle training data with missing attributes. It allows attribute values to be marked as '?' for missing values. Missing attribute values are simply not used in gain and entropy calculations.

Our DEVDT system consists of three major phases - video processing, data pre-processing and data mining. In the first phase it detects the boundary of video events and also observes some important features of video events. In the second phase, it mainly cleans the data i.e. it eliminates the noise data and reduce the irrelevant data from the original feature and produces as training data. In the final phase, it mines the video events from training data and finally it produces pure and relevant events. In the data mining, the information gain ratio criterion is used to determine the most appropriate attribute for extraction due to its efficiency and simplicity for decision tree generation process.

In our future work, this framework will be tested and extended in various types of video events like movies, news, traffic videos (raw video events) and medical video events like ultra sound videos.

ACKNOWLEDGMENT

We would like to thank everyone, who has motivated and supported us for preparing this Manuscript.

REFERENCES

- [1] Simeon J. Simoff, Chabane Djeraba and Osmar R. Zaiane, "MDM/KDD2002: Multimedia Data Mining between Promises and Problems", SIGKDD Explorations, 2002.
- [2] Shroff, N, Turaga,P, and Chellappa,R, Video Prices: Highlighting Diverse Aspects of Videos, Multimedia, IEEE Transactions, and December 2010.
- [3] J.R. Quinlan. **C4.5**: Programs for Machine Learning. Morgan Kaufmann, San Mateo, CA, 1993.
- [4] X. Zhu, X. Wu, A. K. Elmagarmid, Z. Feng, and L. Wu, "Video data mining: Semantic indexing and event detection from the association perspective," IEEE Trans.2005.
- [5] Dianhui Wang, Yong-Soo Kim, Seok Cheon Park, Chul Soo Lee and Yoon Kyung Han, "Learning Based Neural Similarity Metrics for Multimedia Data Mining" Soft Computing, February 2007.
- [6] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," Proceedings of the 1994 International Conference on Very Large Databases, Santiago, Chile, September 1994.
- [7] V. Gudivada, V.V. Raghavan, and K. Vanapipat, 'A Unified Approach to Data Modeling and Retrieval for a Class of Image Database Applications,' In Multimedia Database Systems, V.S. Subrahmanian and S. Jajodia (Eds.), Springer-Verlag, Berlin, Germany, 1996.

- [8] Benoit Huet, Alan Smeaton, Ketan Mayer-Patel and Yannis Avrithis "Advances in Multimedia Modeling"Proc. of International Multimedia Modeling Conference, MMM 2009.
- [9] W.I. Grosky, F. Fotouhi, and Z. Jiang, "Using Metadata for the Intelligent Browsing of Structured Media Objects," In Managing Multimedia Data: Using Metadata to Integrate and Apply Digital Data A. Sheth and W. Klas (Eds.), McGraw Hill Publishing Company, New York, 1998.
- [10] W.I. Grosky, "Managing Multimedia Information in Database Systems", Communications of the ACM, Volume 40, Number 12 (December 1997).
- [11] Michael N; Large database decision tree classifiers; Decision Trees & Data Mining, March 2006.
- [12] S.C. Chen, M.L. Shy, C. Zhang, L. Luo, and M. Chen, "Detection of Soccer Goal Events using Joint Multimedia Features and Classification Rules", Proc. of International Workshop on Multimedia Data Mining (MDWKDD '2003).
- [13] Rosenfeld A., D. Doermann, D. DeMenthon, Eds., Video Mining, Kluwer, 2003.
- [14] J. Assfalg, M. Bertini, A. Del Bimbo, W. Nunziati, and P. Pala, "Soccer highlights detection and recognition using HMMs", proc. of IEEE International Conference on multimedia and Expo,2002.
- [15] Deng,H.; Runger, G.; Tuv, E. "Bias of importance measures for multi-valued attributes and solutions". Proc. of the 21st International Conference on Artificial Neural Networks (ICANN), 2011.
- [16] S.B. Kotsiantis; Supervised Machine Learning: A Review of Classification Techniques, Informatica 31(2007), July 2007.
- [17] Shu-Ching Chen Mei-Ling Shyu Min Chen, Chengcui Zhang. A Decision Tree-based Multimodal Data Mining Framework for Soccer Goal Detection.

AUTHORS PROFILE

Dr P. Premchand is a professor in department of Computer Science & Engineering, Osmania University, Hyderabad, A.P and India. He completed his ME (Computer Science) from Andhra University, A.P. He has received Ph.D degree from Andhra University, A.P. He guided many scholars towards the award of Ph.D degree from various Universities. He was a Director of AICTE, New Delhi, during 1998-99. He also worked as Head of the Dept of CSE and Additional Controller of Examinations, Osmania University, AP. Now currently he is a chairman of BOS, Faculty of Engineering, and O.U. from 2007 to tilldate.

Prof Shaik.Abdul Nabi is the Head of the Dept. of Computer Science & Engineering, AVN Inst.Of Engg.Tech, Hyderabad, AP, and India. He completed his B.E (Computer Science) from Osmania University, A.P. He received his M.Tech. From JNTU College of Engg.,Hyderabad and currently pursuing Ph.D. in the area of Data Mining from Acharya Nagarjuna University, Guntur, AP and India. He is a certified professional by Microsoft. His expertise areas are Data warehousing and Data Mining, Data Structures & UNIX Networking Programming.

Mr.Shaik Rasool is working as a Asst.Prof in Dept of Computer Science &Engineering in S.C.E.T, Hyderabad, India. He received the Bachelor of Technology in Computer Science & Engineering from Jawaharlal Nehru Technological University, Hyderabad, India in 2008. He is completed Master of Technology in Computer Science & Engineering from S.C.E.T., Hyderabad, India. His main research interest includes Data mining, Network Security, Information Security, cloud computing.

An Approach to Improve the Representation of the User Model in the Web-Based Systems

Yasser A. Nada

Computer Science Dept., College of Computer Science
and Information Technology,
Taif University, Kingdom of Saudi Arabia.

Khaled M. Fouad

Computer Science Dept., College of Community,
Taif University, Kingdom of Saudi Arabia

Abstract—A major shortcoming of content-based approaches exists in the representation of the user model. Content-based approaches often employ term vectors to represent each user's interest. In doing so, they ignore the semantic relations between terms of the vector space model in which indexed terms are not orthogonal and often have semantic relatedness between one another.

In this paper, we improve the representation of a user model during building user model in content-based approaches by performing these steps. First is the domain concept filtering in which concepts and items of interests are compared to the domain ontology to check the relevant items to our domain using ontology based semantic similarity. Second, is incorporating semantic content into the term vectors. We use word definitions and relations provided by WordNet to perform word sense disambiguation and employ domain-specific concepts as category labels for the semantically enhanced user models. The implicit information pertaining to the user behavior was extracted from click stream data or web usage sessions captured within the web server logs.

Also, our proposed approach aims to update user model, we should analysis user's history query keywords. For a certain keyword, we extract the words which have the semantic relationships with the keyword and add them into the user interest model as nodes according to semantic relationships in the WordNet.

Keywords—User model; Domain ontology; Semantic Similarity; Wordnet.

I. INTRODUCTION

User model [1] is a collection of personal information. The information is stored without adding further description or interpreting this information. It is comparable to a getting-setting mechanism of classes in object-oriented programming, where different parameters are set or retrieved. User model represents cognitive skills, intellectual abilities, intentions, learning styles, preferences and interactions with the system. These properties are stored after assigning them values. These values may be final or change over time.

The Semantic Web [2] “transforms the Web by providing machine understandable and meaningful descriptions of Web resources”. Making the Web content machine understandable, allowing agents and applications to access a variety of heterogeneous resources, processing and integrating the content, and producing added value for the user. Data on the

Web must be defined and linked in a way that can be used for more effective discovery, automation, integration, and reuse across various applications.

The personalization aspects [3] of the user interests or profiles can form a good representation of the learning context, which promises to enhance the usage of learning content. The key knowledge nugget in any personalization strategy for e-learning is an accurate user model. User Modeling is an active research area in e-learning and personalization, especially when abstracting the user away from the problem an abstraction that has, over the years, contributed to the design of more effective e-learning systems. Despite this improvement, the main focus in most systems, for the past decade, has been on models that are “good for all users”, and not for a specific user.

Our proposed approach is to propose improvements in the representation of a user model during building user model in content-based approaches by performing the next steps. First step is domain concept filtering in which concepts and items of interests are compared to the domain ontology to check the relevant items to our learning domain using ontology based semantic similarity. Second step is incorporating semantic content into the term vectors. We use word definitions and relations provided by WordNet to perform word sense disambiguation and employ domain-specific concepts as category labels for the semantically enhanced user models. The implicit information pertaining to the user behavior was extracted from click stream data or web usage sessions captured within the web server logs. The method of representing semantic user model was proposed in [4].

Also, our proposed approach update user model, we should analysis learner's history query keywords. For a certain keyword, we extract the words which have the semantic relationships with the keyword and add them into the user interest model as nodes according to semantic relationships in WordNet. The method of updating user model was proposed in [5, 6].

II. RELATED WORKS

In [7], authors proposed an idea of adaptation using semantic web techniques with reduced cost of user profile acquisition. Cost-effectiveness is achieved by use of distributed hash tables allowing effective store and lookup operation. Actually DHT operations have to be based on unique IDs which can be easily transformed into keys by means of hash

function employed in particular DHT implementation. Such approach is acceptable for rule based adaptation systems which do not require information about similarity amongst user profiles to decide.

A method was proposed [8] for creating hierarchical user profiles using Wikipedia concepts as the vocabulary for describing user interests. Authors proposed a method for distinguishing informational and recreational interests in the profile from the commercial interests. They developed ways of mapping documents to Wikipedia concepts for the purpose of profile generation.

It was presented [9] a framework for content-based retrieval integrating a relevance feedback method with a word sense disambiguation (WSD) strategy based on WordNet for inducing semantic user profiles. Hypothesis of authors is that substituting words with synsets produces a more accurate document representation that could be successfully used by learning algorithms to infer more accurate user profiles. These semantic profiles will contain references to concepts defined in lexicons or ontologies.

In paper [10], authors combines the ontology and concept space, indicates the feature items of user profile with semantic concepts, calculates learner's interest-level to the topic through establishing the word frequency and utilize the suitable calculation methods, mining the concepts within the user's feedback files and the relationship between concepts, combines user's short-term interests and long-term interests to create user profiles model with semantic concept hierarchy tree and embody the drifting of user profile and improves and completes the user profiles model consistently on the related feedback mechanism.

Authors have proposed [11, 12] an approach to personalized query expansion based on a semantic user model. They discussed the representation and construction of the user model which represents individual user's interests by semantic mining from user's resource searching process in order to perceive the semantic relationships between user's interests which are barely considered in traditional user models and to satisfy the requirement of providing personalized service to users in e-Learning systems.

It has been described in [13] a personalized search approach that represents the user profile as a weighted graph of semantically related concepts of predefined ontology, namely the ODP (<http://www.dmoz.org>). The user profile is built by accumulating graph based query profiles in the same search session. We define also a session boundary recognition mechanism that allows using the appropriate user profile to re-rank search results of queries allocated in the same search session.

III. THE PROPOSED APPROACH

A user model is an internal representation of the user's properties. Before a user model can be used it has to be constructed. This process requires many efforts to gather the required information and finally generate a model of the user. The effectiveness of a user profile depends on the information the system delivers to the user. If a large proportion of information is irrelevant, then the system becomes more of an

annoyance than a help. This problem can be seen from another point of view; if the system requires a large degree of customization, then the user will not be willing to use it anymore.

Depending on the content and the amount of information about the user, which is stored in the user profile, a user can be modeled. Thus, the user profile is used to retrieve the needed information to build up a model of the user. The behavior of an adaptive system varies according to the data from the user model and the user profile. Without knowing anything about the user, a system would perform in exactly the same way for all users [1]. Representation of user model [14, 15] is a necessary factor for building effective and accurate adaptive systems. Adaptive systems compare user profiles to some reference profiles or item characteristics in order to predict the user's model in considering items. The outcome of that process depends on the ability to accurately identify and represent the user's model.

The presented approach for constructing a semantically enhanced user model that represents the user's interests from web-log data [16] (web usage logs). The goal of incorporating the semantic content of the web pages to build the semantically enhanced user models is to address the high dimensionality problem and semantic inadequacy of the Vector Space Model [17, 18, 19] on which the initial user model was based, and to map conceptually related terms. To enrich the user model during the user is browsing the pages and navigate the web-based system the user model must be updated. To update user model our proposed approach analyzes user's history query keywords by using WordNet.

To acquire user interests, we must extract the user behavior and visited page address from web-log data. Then we analyze the visited pages to acquire the terms in the pages that can be considered as concepts in the user model. The extracted terms are represented by Vector Space Model [17, 18, 19] that is adapted to our proposed system to achieve effective representations of documents where each document is identified by an n-dimensional feature vector for each dimension corresponds to a distinct term. Each term in a given document vector has an associated weight.

The term vector serves as the initial term-based user model (IUM) upon which we intended to improve. To build a semantically enhanced user model (SUM), we used refined domain-specific concepts. First we obtained a list of domain-specific concepts from domain ontology. Then we performed term-to concept mapping between terms in the initial user model (IT-UM) and domain related concepts based on concept hierarchies in WordNet. The final product is a semantically enhanced user model (SUM) in which terms are mapped to related high-level concepts.

The semantic User Model (SUM) can be updated using user query and WordNet. For a certain keyword, we extract the words which have the semantic relationships with the keyword and add them into the learner interest model as nodes according to semantic relationships in WordNet.

The goal of incorporating the semantic content of the web pages to build the semantically enhanced user models was to

address the high dimensionality problem and semantic inadequacy of the vector space model, on which the initial user model was based, and to map conceptually related terms. To enrich the user model during the user is browsing the pages and navigate the web-based system the user model must be updated. To update user model our proposed approach analyzes user's history query keywords by using WordNet.

The proposed approach architecture is shown in figure 1.

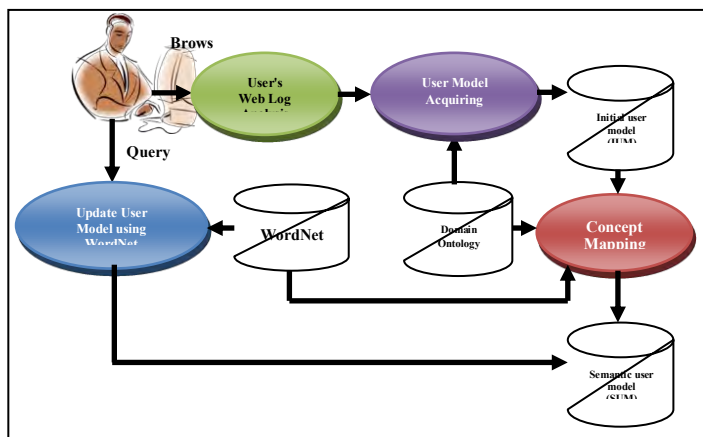


Figure 1: Proposed System Architecture

A. User's Web Log Analysis

Web usage mining [20], the process of discovering patterns from web data using data mining methods, strives to find learner preferences based on the web-logs that reside on servers. Web log [16] records each transaction, which was executed by the browser at each web access. Each line in the log represents a record with the IP address, time and date of the visit, accessed object and referenced object. In such data, we follow sequences in visiting individual pages by the learner, who is, under certain condition, identified by the IP address. In sequences, we can look for learners behavior patterns.

The data from Web logs, in its raw form, is not suitable for the application of usage mining algorithms. The data need to be cleaned and preprocessed. To perform log data analysis, the data pre-processing process must be accomplished. The data pre-processing is the process of cleaning and transforming raw data sets into a form suitable for web mining. The task of the data pre-processing module is therefore, to obtain usable datasets from raw web log files, which, in most cases, contain a considerable amount of incomplete and irrelevant information.

The overall data preparation process [21, 22] is briefly described in figure 2.

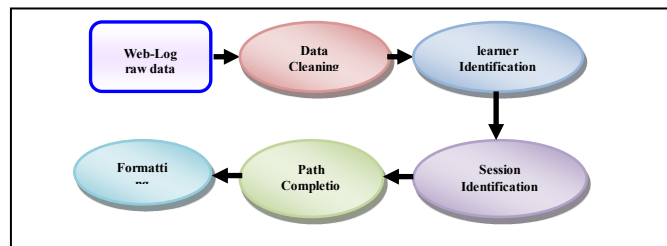


Figure 2: data preparation process for web log

Data Cleaning: to remove accesses to irrelevant items (such as button images), accesses by Web crawlers (i.e. non-human accesses), and failed requests.

Learner Identification: Because web logs are recorded in a sequential manner as they arrive, therefore, records for a specific learner are not necessary recorded in consecutive order rather they could be separated by records from other learners.

Session Identification: To divide pages accessed by each learner into individual sessions. A session is a sequence of pages visited by a learner. We also call it as a usage sequence.

Path Completion: To determine if there are important accesses which are not recorded in the access log due to caching on several levels.

Formatting: Format the data to be readable by data mining systems.

Once web logs are preprocessed, useful web usage patterns may be generated by applying data mining techniques. Table 1 shows a sample of web log data after preprocessing process.

TABLE I. SAMPLE OF WEB LOG DATA

Visit Time	UserId	URL
20090405202122	10	http://www.cs.bu.edu/teaching/
20090405203225	19	http://www.cs.bu.edu/teaching/unix/intro/
20090406081905	10	http://www.cs.bu.edu/teaching/cs113/spring-2000/object/
20090407091215	11	http://www.aw-bc.com/brookshear/
20090407082621	19	http://hortle.ccsu.edu/java5/Notes/chap21/ch21_1.html

The outputs of this step are web based learning materials; that the learner explored and preferred it, and the behavior pattern of the learner. The learner behavior is used to acquiring knowledge requirement for learners based on course ontology.

B. Domain ontology developing based knowledge engineering approach

Ontology engineering is a subfield of knowledge engineering that studies the methods and methodologies for building ontologies. It researches the ontology development process, the ontology life cycle, the methods and methodologies for building ontologies, and the tools suite and languages that support them. Knowledge Engineering field usually uses the IEEE 1074-2006 standard [23] as reference criteria. The IEEE 1074-2006 is a standard for developing a software project life cycle processes. It describes the software development process, the activities to be carried out, and techniques that can be used for developing software.

It was proposed [24] a knowledge engineering approach to build domain ontology. Figure 3 shows main steps of the ontology development process.

Identify the purpose and requirement specification concerns to clear identify the ontology purpose, scope and its intended use, that is the competence of the ontology. Ontology acquisition is to capture the domain concepts based on the ontology competence. The relevant domain entities (e.g. concepts, relations, slots, and role) should be identified and organized into hierarchy structure. This phase involves three steps as follows: first, enumerate important concepts and terms in this domain; second, define concepts, properties and

relations of concepts, and organize them into hierarchy structure; third, consider reusing existing ontology.

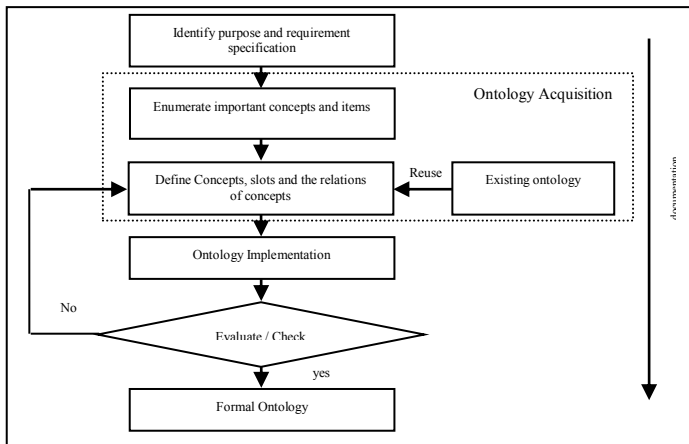


Figure 3 Main steps of the ontology development

Ontology implementation aims to explicit represent the conceptualization captured in a formal language. Evaluation/Check means that the ontology must be evaluated to check whether it satisfies the specification requirements. Documentation means that all the ontology development must be documented, including purposes, requirements, textual descriptions of the conceptualization, and the formal ontology.

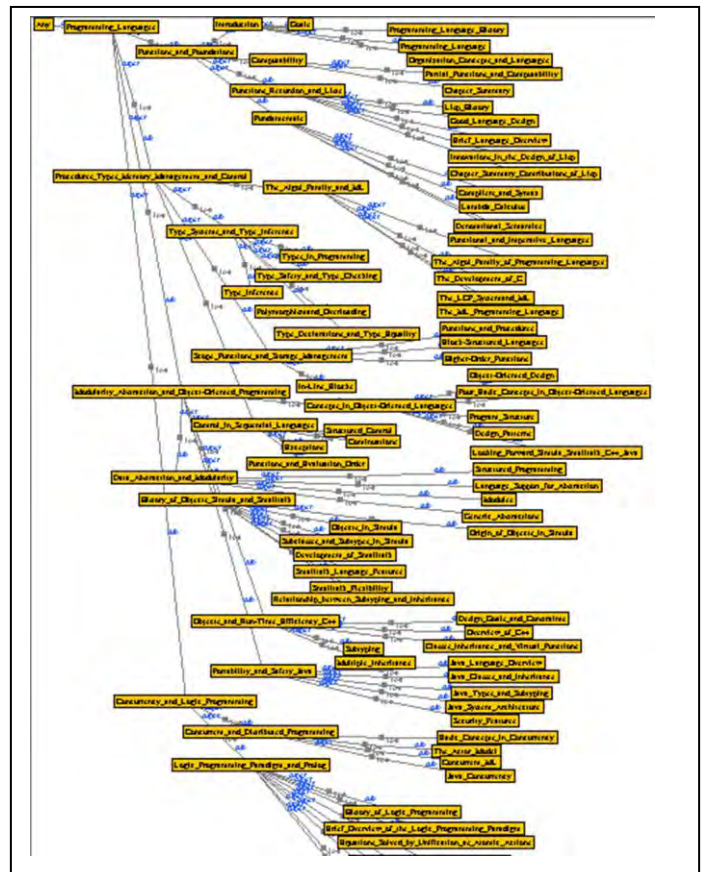
Our domain focuses "programming languages" course. We use Hozo [25] as our ontology editor. Since Hozo is based on an ontological theory of a role-concept, it can distinguish concepts dependent on particular contexts from so-called basic concepts and contribute to building reusable ontologies. A role-concept [24] represents a role which an object plays in a specific context and it is defined with other concepts. On the other hand, a basic-concept does not need other concepts for being defined. An entity of the basic concept that plays a role-concept is called a role-holder. Figure 4 shows part of our domain ontology and the extracted OWL [26] is shown in figure 5.

C. User Model Acquiring

In the proposed system [6, 22], user interest model's knowledge expression uses the thought, which is based on the space vector model's expression method and the domain ontology. This method acquires user's interest was shown in [6, 22]. Figure 5 shows certain steps to acquire user interest.

D. Document Representation

The Vector Space Model [27, 28] is adapted in our proposed system to achieve effective representations of documents. Each document is identified by n-dimensional feature vector for each dimension corresponds to a distinct term. Each term in a given document vector has an associated weight.



```

- <owl:Class rdf:ID="Programming_Languages">
  <rdfs:label>Programming_Languages</rdfs:label>
</owl:Class>
- <owl:Class rdf:ID="Functions_and_Foundations">
  <rdfs:label>Functions_and_Foundations</rdfs:label>
  <rdfs:subClassOf rdf:resource="#Programming_Languages" />
</owl:Class>
- <owl:Class rdf:ID="Introduction">
  <rdfs:label>Introduction</rdfs:label>
  <rdfs:subClassOf rdf:resource="#Functions_and_Foundations" />
</owl:Class>
- <owl:Class rdf:ID="Programming_Language">
  <rdfs:label>Programming_Language</rdfs:label>
  <rdfs:subClassOf rdf:resource="#Introduction" />
</owl:Class>
- <owl:Class rdf:ID="Goals">
  <rdfs:label>Goals</rdfs:label>
  <rdfs:subClassOf rdf:resource="#Introduction" />
</owl:Class>
- <owl:Class rdf:ID="Programming_Language_History">
  <rdfs:label>Programming_Language_History</rdfs:label>
  <rdfs:subClassOf rdf:resource="#Introduction" />
</owl:Class>
- <owl:Class rdf:ID="Organization_Concepts_and_Languages">
  <rdfs:label>Organization_Concepts_and_Languages</rdfs:label>
  <rdfs:subClassOf rdf:resource="#Introduction" />
</owl:Class>
- <owl:Class rdf:ID="Computability">
  <rdfs:label>Computability</rdfs:label>
  <rdfs:subClassOf rdf:resource="#Functions_and_Foundations" />
</owl:Class>
- <owl:Class rdf:ID="Partial_Functions_and_Computability">
  <rdfs:label>Partial_Functions_and_Computability</rdfs:label>
  <rdfs:subClassOf rdf:resource="#Computability" />
</owl:Class>
- <owl:Class rdf:ID="Chapter_Summary">
  <rdfs:label>Chapter_Summary</rdfs:label>
  <rdfs:subClassOf rdf:resource="#Computability" />
</owl:Class>
  
```

Figure 5: Extracted OWL for the Domain Ontology

TABLE III SHOWS THE TERM WEIGHTS IN DIFFERENT DOCUMENTS

Items / DOC	Programming Language	Programing of Lists	Functions and Foundations	Programming of Recursion	Procedures Types	Memory Management and Control	Object Oriented Programming	Structured Programming	Concurreny and Logic Programing	Distributed Programing	Logic Programming	ML Programming Language
Doc1	0.3899	0.8490	0.4622	0.8490	0.5034	0.0000	0.7320	0.8390	0.5034	0.0000	0.5034	0.1830
Doc 2	0.2599	0.2830	0.0000	0.7075	0.6712	0.1830	0.0000	0.0000	0.0000	1.3156	0.3356	0.7320
Doc 3	0.5198	0.8490	0.7703	0.5660	0.0000	0.7320	0.5490	0.8390	0.6712	0.0000	0.0000	0.0000
Doc 4	0.2599	0.5660	0.6162	0.0000	0.0000	0.9150	0.7320	0.8390	0.0000	0.0000	0.0000	0.0000
Doc 5	0.6498	0.9905	0.7703	0.0000	0.5034	0.0000	0.3660	0.8390	1.3425	0.8771	0.0000	0.5490
Doc 6	0.2599	0.8490	0.0000	0.7075	0.8390	0.0000	0.0000	0.0000	0.6712	0.4385	0.5034	1.4641
Doc 7	0.1300	0.7075	0.7703	0.0000	0.0000	0.9150	1.2811	0.8390	1.1747	0.2193	0.0000	0.0000
Doc 8	0.0000	0.0000	0.0000	0.7075	0.0000	0.0000	0.0000	0.5034	1.0068	1.0963	0.6712	0.0000
Doc 9	0.2599	0.7075	0.6162	0.1415	0.0000	0.7320	0.9150	0.8390	0.3356	0.0000	0.1678	0.0000
Doc 10	0.6498	0.0000	0.7703	0.8490	0.0000	1.4641	0.0000	0.0000	0.0000	0.0000	1.0068	0.0000
Doc 11	0.9097	0.0000	0.0000	0.4245	0.5034	0.0000	1.4641	0.1678	0.6712	1.7541	0.5034	0.5490
Doc 12	0.5198	0.0000	0.0000	0.0000	0.6712	0.5490	0.0000	1.1747	1.0068	1.5349	0.0000	0.7320
Doc 13	0.2599	0.4245	0.7703	0.2830	0.3356	0.0000	0.0000	0.8390	0.0000	0.0000	0.3356	0.0000
Doc 14	0.0000	0.2830	0.7703	0.1415	0.1678	0.7320	0.1830	0.0000	0.0000	0.0000	0.1678	0.1830
Doc 15	0.6498	0.8490	1.0784	0.7075	1.1747	0.9150	0.9150	0.0000	0.0000	0.0000	0.8390	0.7320
Doc 16	0.0000	0.5660	0.4622	0.5660	1.0068	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0980

TABLE II SHOWS THE TERM FREQUENCY IN DIFFERENT DOCUMENTS

Items / DOC	Programming Language	Programing of Lists	Functions and Foundations	Programming of Recursion	Procedures Types	Memory Management and Control	Object Oriented Programming	Structured Programming	Concurreny and Logic Programing	Distributed Programing	Logic Programming	ML Programming Language
Doc1	3	6	3	6	3	0	4	5	3	0	3	1
Doc 2	2	2	0	5	4	1	0	0	0	6	2	4
Doc 3	4	6	5	4	0	4	3	5	4	0	0	0
Doc 4	2	4	4	0	0	5	4	5	0	0	0	0
Doc 5	5	7	5	0	3	0	2	5	8	4	0	3
Doc 6	2	6	0	5	5	0	0	0	4	2	3	8
Doc 7	1	5	5	0	0	5	7	5	7	1	0	0
Doc 8	0	0	0	5	0	0	0	3	6	5	4	0
Doc 9	2	5	4	1	0	4	5	5	2	0	1	0
Doc 10	5	0	5	6	0	8	0	0	0	0	6	0
Doc 11	7	0	0	3	3	0	8	1	4	8	3	3
Doc 12	4	0	0	0	4	3	0	7	6	7	0	4
Doc 13	2	3	5	2	2	0	0	5	0	0	2	0
Doc 14	0	2	5	1	1	4	1	0	0	0	1	1
Doc 15	5	6	7	5	7	5	5	0	0	0	5	4
Doc 16	0	4	3	4	6	0	0	0	0	0	0	6
No of Doc's	13	12	11	12	10	9	9	10	10	7	10	9

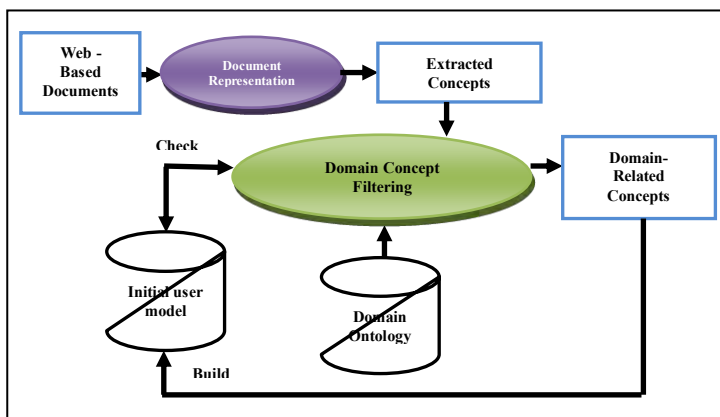


Figure 5: steps to acquire learner interest

The weight is a function of the term frequency, collection frequency and normalization factors. Different weighting approaches may be applied by varying this function. Hence, a document j is represented by the document vector d_j :

$d_j = (w_{1j}, w_{2j}, \dots, w_{nj})$ Where, w_{kj} is the weight of the k^{th} term in the document j .

The term frequency reflects the importance of term k within a particular document j . The weighting factor may be global or local. The global weighting factors clarify the importance of a term k within the entire collection of documents, whereas a local weighting factor considers the given document only.

The document keywords were extracted by using a term-frequency-inverse-document-frequency (*tf-idf*) calculation [18, 19], which is a well-established technique in information retrieval. The weight of term k in document j is represented as:

$$w_{kj} = tf_{kj} \times (\log_2^n - \log_2^{df_k} + 1)$$

Where: tf_{kj} = the term k frequency in document j , df_k = number of documents in which term k occurs, n = total number of documents in the collection. Table 2 shows the term frequency in different documents.

The main purpose of this step is to extract interested items in the web page, then get term frequency that reflects the importance of the term. Finally, get the weight of terms in the selected page. The output of this step is the weight of terms in selected page that can be used to build learner interest profile. Table 3 shows a sample of the weighted terms in the documents; that found in table 3.

E. Domain Concept Filtering

This process discovers concepts which represent the learner's interests. These concepts and items are compared to the domain ontology to check the relevant items to the learner profile. The most relevant ones update the learner profile. The items relevance is based on ontology-based semantic similarity where browsed items by a learner on the web are compared to the items from a domain ontology and learner profile. The importance is combined with the semantic similarity to obtain a level of relevance. The page items are processed to identify domain-related words to be added to the learner profile. A bag

of browsed items is obtained via a simple word indexing of the page visited by the learner. We filter out irrelevant words using the list of items extracted from domain ontology. Once domain-related items are identified, we evaluate their relevance to learner's interests.

The selected method was used in [29, 30] to compute semantic similarity function (S) based on a domain ontology. The similarity is estimated for each pair of items where one item is taken from a learner profile, while the other one from a set of browsed items.

The functions S_w is the similarity between synonym sets, S_u is the similarity between features, and S_n is the similarity between semantic neighborhoods between entity classes an of ontology p and b of ontology q , and w_w , w_u , and w_n are the respective weights of the similarity of each specification component.

$$S(a^p, b^q) = w_w \times S_w(a^p, b^q) + w_u \times S_u(a^p, b^q) + w_n \times S_n(a^p, b^q)$$

; For $w_w, w_u, w_n \geq 0$;

Weights assigned to S_w , S_u , and S_n depend on the characteristics of the ontologies.

The similarity measures are defined in terms of a matching process [29, 30]:

$$S(a, b) = \frac{|A \cap B|}{|A \cap B| + \alpha(a, b) |A / B| + (1 - \alpha(a, b)) |B / A|}$$

Where A and B are description sets of classes a and b , i.e., synonym sets, sets of distinguishing features and a set of classes in semantic neighborhood; $(A \cap B)$ and (A / B) represent intersection and difference respectively, $||$ is the cardinality of a set; and α is a function that defines relative importance of non-common characteristics. A set of browsed items that are similar to items from the learner profile is considered as a set of items that can be added to this profile.

IV. BUILDING SEMANTIC USER MODEL USING CONCEPT MAPPING

To overcome these weaknesses of term-based representations, an ontology-based representation [33, 34] using wordnet will be performed. Moreover, by defining an ontology base, which is a set of independent concepts that covers the whole ontology, an ontology-based representation allows the system to use fixed-size document vectors, consisting of one component per base concept.

We present a method based on WordNet [35] that improves traditional vector space model. WordNet is an ontology of cross-lexical references whose design was inspired by the current theories of human linguistic memory. English names, verbs, adjectives, and adverbs are organized in sets of synonyms (synsets), representing the underlying lexical concepts. Sets of synonyms are connected by relations. The basic semantic relation between the words in WordNet is synonymy [36]. Synsets are linked by relations such as specific/generic or hypernym /hyponym (is-a), and meronym/holonym (part-whole). The principal semantic relations supported by WordNet is synonymy: the synset

(synonym set), represents a set of words which are interchangeable in a specific context. WordNet [36] consists of over 115,000 concepts (synsets in WordNet) and about 150,000 lexical entries (words in WordNet). This representation requires two more stages: a) the “mapping” of terms into concepts and the choice of the “merging” strategy, and b) the application of a disambiguation strategy.

The purpose of this step is to identify WordNet concepts that correspond to document words [31]. Concept identification is based on the overlap of the local context of the analyzed word with every corresponding WordNet entry. The entry which maximizes the overlap is selected as a possible sense of the analyzed word. The concept identification architecture for the terms in the initial user model is given in figure 6.

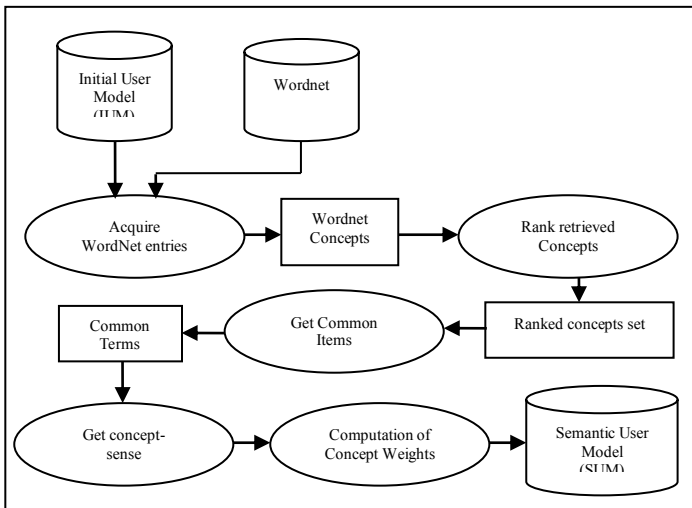


Figure 6: Semantic User Model using Concept Mapping

We use WordNet categories [32] to map all the stemmed words in all documents into their lexical categories. For example, the word “dog” and “cat” both belong to the same category “noun.animal”. Some words also has multiple categories like word “Washington” has 3 categories (noun.location, noun.group, noun.person) because it can be the name of the American president, the city place, or a group in the concept of capital. Some word disambiguation techniques are used to remove the resulting noise added by multiple categories mapping which are disambiguation by context and concept map.

A. The Weight of Concept Computation

The concepts in documents are identified as a set of terms that have identified or synonym relationships, i.e., synsets in the WordNet ontology. Then, the concept frequencies Cf_c are calculated based on term frequency tf_t as follows:

$$Cf_c = \sum_{t_m \in r(c)} tf_{tm}$$

Where $r(c)$ is the set of different terms that belong to concept C . Note that WordNet returns an ordered list of synsets based on a term. The ordering is supposed to reflect how common it is that the term is related to the concept in standard

English language. More common term meanings are listed before less common ones. The authors in [33, 34] have showed that using the first synset as the identified concept for a term can improve the clustering performance more than that of using all the synsets to calculate concept frequencies.

Hypernyms of concepts can represent such concepts up to a certain level of generality. The concept frequencies are updated as follows:

$$hf_c = \sum_{b \in H(c,r)} Cf_b$$

Where $H(c,r)$ is the set of concepts C_H , which are all the concepts within r levels of hypernym concepts of c .

In WordNet, is obtained by gathering all the synsets that are hypernym concepts of synset c within r levels. In particular, $H(c,\infty)$ returns all the hypernym concepts of c and $H(c,0)$ returns just c .

The weight of each concept c in document d is computed as follows:

$$wh_c = hf_c \times idf_c$$

Where idf_c is the inverted document frequency of concept c by counting how many documents in which concept c appears as the weight of each term t in the document d .

The weights of the concepts after mapping the items in table 3 is shown in table 4 after computing the concepts weights.

V. UPDATE USER MODEL USING WORDNET

During the user is working through the web based learning system, user interests' change quite often, and users are reluctant to specify all adjustments and modifications of their intents and interests. Therefore, techniques that leverage implicit approaches for gathering information about users are highly desired to update the user interests that are often not been fixed.

In order to update user interest [6, 37], first of all, we should analysis user's history query keywords. For a certain keyword, we extract the words which have the semantic relationships with the keyword and add them into the user interest model as nodes according to semantic relationships in WordNet.

With new words added constantly, user is always interested in the kind of the words with a higher score which standard for some type of knowledge. We must constantly, update the user model after the users enter the new specific keywords. User model is updated by the new keywords. The incremental updating strategy is used here, and gives the related words the different score according to the relations which reflect their importance of different words in order to render the interestingness of the words. As a result, the words that are more frequent have a higher score. Because of history keywords have the order, the keywords which are inquired later always have more meaning than the keywords which are inquired earlier; it need multiply a factor of attenuation β when increasing the score. Because the keywords are added

constantly and the scale of the user model becomes bigger, some old nodes must be removed in order to reduce user interest model.

The main steps of this method can be described as follows:

1) *If a new keyword is found in the original user model, we increase the score of the related nodes directly. That is, the node is given by five score after multiplying a factor of attenuation β . If it is not found, we must create a new word node and give it five score.*

2) *Finding the following three relations between new keywords and inputted words in the user model based on the WordNet:*

a) *Synonymous relations: obtain the synonym set and insert every synonym into the original user model in turn. If the synonym is found in the original user model, we increase the score of the related nodes directly. That is, the node is given by four score after multiplying a factor of attenuation β . Otherwise, create a new word node with four score and add a new undirected edge labeled synonym relation.*

b) *Hyponym or Hypernym relations: obtain the hyponym or hypernym set and insert every word into the original user model in turn. If the word is found in the original user model, we increase the score of the related nodes directly. That is, the node is given by two score after multiplying a factor of attenuation β . Otherwise, create a new word node with two score and add a new directed edge labeled hyponym or hypernym relation.*

c) *Meronym or Holonym relations: obtain the meronym or holonym set and insert every word into the original user model in turn. If the word is found in the original user model, we increase the score of the related nodes directly. That is, the node is given by one score after multiplying a factor of attenuation β . Otherwise, create a new word node with one score and add a new directed edge labeled meronym or holonym relation.*

d) *In order to reduce user interest model, the nodes which have the lower score must be removed after some time.*

CONCLUSION

We have presented in this paper a novel approach for conceptual document indexing. Our contribution concerns two main aspects. The first one consists on a concept-representation approach of the initial user model items based on the use of WordNet. The approach is not new but, we proposed new techniques to identify concepts and to weight them. In addition to the semantic representation approach to build the semantic user model, we proposed approach to update the user model using the Wordnet.

REFERENCES

- [1] F. Christoph, (2005). User Modeling and User Profiling in Adaptive E-learning Systems, Master's Thesis At Graz University of Technology.
- [2] P. Apple Wai and S. Horace.(2007), Educational Ontologies Construction for Personalized Learning on the Web, Studies in Computational Intelligence (SCI) 62, 47–82, Springer-Verlag Berlin Heidelberg.
- [3] Z. Leyla, R. Elizabeth, W. Robert. (2009), The Effectiveness of Personalization in Delivering E-learning Classes, 2009 Second International Conferences on Advances in Computer-Human Interactions, 978-0-7695-3529-6/09, IEEE.
- [4] A. Palakorn, H. Hyoil, N. Olfà, J. Roberta. (2007), Semantically Enhanced User Modeling, SAC'07, March 11-15, 2007, Seoul, Korea., ACM 1-59593-480-4/07/0003.
- [5] H. Changqin, J. Ying, D. Rulin. (2010), A semantic web-based personalized learning service supported by on-line course resources, Networked Computing (INC), 2010 6th International Conference, Print ISBN: 978-1-4244-6986-4, IEEE.
- [6] H. Hany, F. Khaled. (2010). Semantic web based Approach to learn and update Learner Profile in Adaptive E-Learning, Al-Azhar Engineering Eleventh International Conference, December 23-26.
- [7] K. Tom, J. Ivan. (2008), Semantic User Profile Acquisition and Sharing, International Conference on Computer Systems and Technologies – CompSysTech'08, Gabrovo, Bulgaria, June 12-13, 2008 , ACM 2008 ISBN: 978-954-9641-52-3/08/06.
- [8] R. Krishnan and K. Komal. (2009), Creating User Profiles Using Wikipedia, A.H.F. Laender et al. (Eds.): ER 2009, LNCS 5829, pp. 415–427, 2009. Springer-Verlag Berlin Heidelberg.
- [9] M. Degemmis, P. Lops, and G. Semeraro. (2006), WordNet-Based Word Sense Disambiguation for Learning User Profiles, EWMF/KDO 2005, LNAI 4289, pp. 18–33, Springer-Verlag Berlin Heidelberg.
- [10] W. Cuncun. Chongben , T. Hengsong. (2009), A Personalized Model for Ontology-driven User Profiles Mining, 2009 International Symposium on Intelligent Ubiquitous Computing and Education, DOI 10.1109/IUCE.2009.128, 978-0-7695-3619-4/09, IEEE.
- [11] L. Xiaojian, C. Shihong. (2009), Personalized Query Expansion Based on Semantic User Model in e-Learning System, 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery, DOI 10.1109/FSKD.2009.787, 978-0-7695-3735-1/09, IEEE.
- [12] L. Xiaojian, C. Shihong. (2009), Research on Personalized User Model Based on Semantic Mining from Educational Resources Searching Process, 2009 International Joint Conference on Artificial Intelligence, DOI 10.1109/JCAI.2009.67, 978-0-7695-3615-6/09, IEEE.
- [13] D. Mariam, T. Lynda, B. Mohand. (2009), A Session Based Personalized Search Using An Ontological User Profile, SAC'09 March 8-12, 2009, Honolulu, Hawaii, U.S.A., ACM 978-1-60558-166-8/09/03.
- [14] Z. Ning, L. Yuefeng. (2003). Ontology-Based Web Mining Model: Representation of User Profiles. Proceeding of the IEEE/WIC International Conference on Web Intelligence (WI'03), 0-7695-1932-6/03, IEEE.
- [15] S. Ahu, M. Bamshad, B. Robin. (2007). Ontological User Profiles for Representing Context in Web Search. 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology – Workshops. 0-7695-3028-1, IEEE.
- [16] <http://www.w3.org/TR/WD-logfile.html>.
- [17] F. Christoph. (2005), User Modeling and User Profiling in Adaptive E-learning Systems, Master's Thesis at Graz University of Technology, Copyright 2005 by Christoph Fröschl.
- [18] L. Xiaojian, C. Shihong. (2009), Personalized Query Expansion Based on Semantic User Model in e-Learning System, 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery, DOI 10.1109/FSKD.2009.787, 978-0-7695-3735-1/09, IEEE.
- [19] A. Palakorn, H. Hyoil, N. Olfà, J. Roberta. (2007), Semantically Enhanced User Modeling, SAC'07, March 11-15, 2007, Seoul, Korea., ACM 1-59593-480-4/07/0003.
- [20] G. PAOLO, F. SILVIA. (2009). Applied Data Mining for Business and Industry Second Edition, Chapter 6: Describing website visitors, ISBN: 978-0-470-05887-9 (Pbk), John Wiley.
- [21] K. Natheer, C. Chien-Chung. (2005). Web Usage Mining Using Rough Sets, NAFIPS 2005 - 2005 Annual Meeting of the North American Fuzzy Information Processing Society, 0-7803-9187-X/05, IEEE.
- [22] F. Khaled, M. Hany, M. Nagdy. (2011). Semantic Web supporting Adaptive E-Learning to build and represent Learner Model. The Second International Conference of E-learning and Distance Education (eLI 2011) – Riyadh 2011.

- [23] IEEE Standard for Developing Software Life Cycle Processes, IEEE Computer Society, New York (USA), April 26, 1997.
- [24] H. YUN, J. XU, M. J. XIONG. (2009). Development of Domain Ontology for E-learning Course, 978-1-4244-3930-0/09, IEEE.
- [25] K. Kozaki, Y. Kitamura, M. Ikeda, and R. Mizoguchi. (2002). Hozo: An Environment for Building/Using Ontologies Based on a Fundamental Consideration of Role” and “Relationship”, Proc. of the 13th International Conference Knowledge Engineering and Knowledge Management (EKAW2002), Sigüenza, Spain, October 1-4, 2002, pp.213-218.
- [26] <http://www.w3.org/TR/owl-ref/>.
- [27] B. Qiu, W. Zhao. (2009). Student Model in Adaptive Learning System based on Semantic Web, 2009 First International Workshop on Education Technology and Computer Science, 978-0-7695-3557-9/09, IEEE, DOI 10.1109/ETCS.2009.466.
- [28] P. Jianguo, Z. Bofeng, W. Shufeng, W. Gengfeng, and W. Daming. (2007). Ontology Based User Profiling in Personalized Information Service Agent, Seventh International Conference on Computer and Information Technology, 0-7695-2983-6/07, IEEE.
- [29] R. Marek and K. Sayed. (2009). Updating User Profile using Ontology-based Semantic Similarity, FUZZ_IEEE 2009, Korea, August 20-24, 978-1-4244-3597-5, IEEE.
- [30] R. M. Andrea, E. Max J. (2003). Determining Semantic Similarity among Entity Classes from Different Ontologies, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 15, NO. 2, MARCH/APRIL, 1041-4347/03, IEEE.
- [31] B. Fatiha, B. Mohand, T. Lynda, D. Mariam. (2010). Using WordNet for Concept-Based Document Indexing in Information Retrieval, SEMAPRO: The Fourth International Conference on Advances in Semantic Processing, Pages: 151 to 157, IARIA.
- [32] A. Abdelmalek, E. Zakaria, S. Michel, M. Mimoun. (2008). WordNet-based and N-Grams-based Document Clustering: A Comparative Study, Third International Conference on Broadband Communications, Information Technology & Biomedical Applications, 978-0-7695-3453-4/08, IEEE.
- [33] D. Mauro, C. Celia, G. Andrea. (2010). An Ontological Representation of Documents and Queries for Information Retrieval Systems, IEA/AIE 2010, Part II, LNAI 6097, pp. 555–564, Springer-Verlag Berlin Heidelberg.
- [34] C. Celia, G. Andrea. (2006). An Ontology-Based Method for User Model Acquisition. In: Ma, Z. (ed.) Soft computing in ontologies and semantic Web. Studies in fuzziness and soft computing, pp. 211–227., Springer-Verlag Berlin Heidelberg.
- [35] F. Christiane. (2010). WordNet. In R. Poli et al. Theory and Applications of Ontology: Computer Applications, (pp. 231-243). 231-243, DOI: 10.1007/978-90-481-8847-5_10, Springer Science+Business Media B.V.
- [36] A. Abdelmalek Amine, E. Zakaria, and S. Michel. (2010). Evaluation of Text Clustering Methods Using WordNet. The International Arab Journal of Information Technology, Vol. 7, No. 4.
- [37] S. Yan, L. Yun, L. Luan, C. Ling. (2009). A Personalized Search Results Ranking Method Based on WordNet, 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery, 978-0-7695-3735-1/09, IEEE.
- [38] Khaled M. Fouad, Shehab Gamalel-Din, Mofreh A. Hogo, Nagdy M. Nagdy. Adaptive E-Learning System based on Semantic Web and Fuzzy Clustering (IJCSIS).

AUTHORS PROFILE



DR. Yasser A. Nada Was born in Ismailia, Egypt, in 1968. He received the BSc degree in pure Mathematics and Computer Sciences in 1989 and MSc degree for his work in computer science in 2003, all from the Faculty of Science, Suez Canal University, Egypt. In 2007, he received his Ph.D. in Computer Science from the Faculty of Science, Suez Canal University, Egypt. From September 2007 until now, he worked as a lecturer of computer science, Faculty of Computers and Information Systems Taif University, KSA. His research interests include Expert Systems, Artificial Intelligence, Semantic Web Object Oriented Programming, Computer Vision, and Genetic.



Khaled M. Fouad received his Master degree of AI and expert systems. He is currently a PhD candidate in the faculty of engineering AlAzhar University in Egypt. He is working now as lecturer in Taif University in Kingdom of Saudi Arabia (KSA) and is assistant researcher in Central Laboratory of Agriculture Expert Systems (CLAES) in Egypt. His current research interests focus on Semantic Web and Expert Systems.

Solving the MDBCS Problem Using the Metaheuristic— Genetic Algorithm

Genetic algorithm for the MDBCS problem

Milena Bogdanović

University of Niš

Teacher Training Faculty

Partizanska 14, Vranje, Serbia

Abstract— The problems degree-limited graph of nodes considering the weight of the vertex or weight of the edges, with the aim to find the optimal weighted graph in terms of certain restrictions on the degree of the vertices in the subgraph. This class of combinatorial problems was extensively studied because of the implementation and application in network design, connection of networks and routing algorithms. It is likely that solution of MDBCS problem will find its place and application in these areas. The paper is given an ILP model to solve the problem MDBCS, as well as the genetic algorithm, which calculates a good enough solution for the input graph with a greater number of nodes. An important feature of the heuristic algorithms is that can approximate, but still good enough to solve the problems of exponential complexity. However, it should solve the problem heuristic algorithms may not lead to a satisfactory solution, and that for some of the problems, heuristic algorithms give relatively poor results. This is particularly true of problems for which no exact polynomial algorithm complexity. Also, heuristic algorithms are not the same, because some parts of heuristic algorithms differ depending on the situation and problems in which they are used. These parts are usually the objective function (transformation), and their definition significantly affects the efficiency of the algorithm.

By mode of action, genetic algorithms are among the methods directed random search space solutions are looking for a global optimum.

Keywords- graph theory; NP-complete problems; the degree-bounded graphs; Integer linear programming; genetic algorithms.

I. INTRODUCTION

General problems of degree-constrained graphs, consider the nodes of weight or weight on the vertices, where the goal is to find the optimal weighted graph, with set limits for levels of subgraph nodes. This class of combinatorial problems has been extensively studied for use in designing networks. If the input graph is bipartite, ie, if the set of its nodes can be broken down into two nonempty disjoint subsets so that each vertices has one end in each of the two subsets, then these problems are equivalent to the classical transportation problem in terms of operational research. The reason for this extensive study and research issues listed above, lies in their wide application in the areas of networks and routing algorithms.

In the Theory of complexity, NP (nondeterministic polynomial time) is a set of decision problems that can be

solved by nondeterministic Turing machine. The importance of this class of decision problems is that it contains many interesting problems of search and optimization, where we want to determine whether there is some solution to the problem, but whether this is the optimal solution. Therefore, the challenge with NP problem is to find the answer in an efficient manner, as an effective way to verify the response, ie. solution already exists. Since many important problems in this class, intensive efforts were invested to find in polynomial time algorithms for solving problems in class NP. However, a large number of NP problems has resisted these efforts, and apparently, they require time polynomial is not even close! Are these problems really are not solvable in polynomial time is one of the biggest open questions in computer science.

The easiest way to prove that an problem a NP-complete problem is to first prove that the NP and then to an already known NP-complete problem down to him. It is therefore useful to know the various NP-complete problems.

The problems of class NP-complete are classified into the following groups:

- 1) *Covering and Partitioning*
- 2) *Subgraphs and Supergraphs*
- 3) *Vertex Ordering*
- 4) *Iso- and Other Morphisms*
- 5) *Miscellaneous*

From the group Subgraph and supergraphs distinguishes the following NP-problems:

- 1) *MAXIMUM INDEPENDENT SET*
- 2) *MAXIMUM INDEPENDENT SEQUENCE*
- 3) *MAXIMUM INDUCED SUBGRAPH WITH PROPERTY P*
- 4) *MINIMUM VERTEX DELETION TO OBTAIN SUBGRAPH WITH PROPERTY P*
- 5) *MINIMUM EDGE DELETION TO OBTAIN SUBGRAPH WITH PROPERTY P*
- 6) *MAXIMUM INDUCED CONNECTED SUBGRAPH WITH PROPERTY P*
- 7) *MINIMUM VERTEX DELETION TO OBTAIN CONNECTED SUBGRAPH WITH PROPERTY P*

- 8) MAXIMUM DEGREE-BOUNDED CONNECTED SUBGRAPH
- 9) MAXIMUM PLANAR SUBGRAPH
- 10) MINIMUM EDGE DELETION K-PARTITION
- 11) MAXIMUM K-COLORABLE SUBGRAPH
- 12) MAXIMUM SUBFOREST
- 13) MAXIMUM EDGE SUBGRAPH
- 14) MINIMUM EDGE K-SPANNER
- 15) MAXIMUM K-COLORABLE INDUCED SUBGRAPH
- 16) MINIMUM EQUIVALENT DIGRAPH
- 17) MINIMUM INTERVAL GRAPH COMPLETION
- 18) MINIMUM CHORDAL GRAPH COMPLETION

The problem MDBCS is stated as follows:

INPUT: a graph $G = (V, E)$, the function of weight on the vertices (or weighting function) $w: E \rightarrow \mathbb{R}^+$, and an integer $d \geq 2$.

SOLUTION: a subset of $E' \subseteq E$ such that the subgraph $G' = (V, E')$ is connected and that there is no node with degree exceeding d .

MEASUREMENTS: Total weight of found subgraph, i.e. $\sum_{e \in E'} w(e)$.

Considering an undirected graph $G = (V, E)$, where E is the set of vertices, and V is a set of nodes, and let $|V| = n, |E| = m$. Let $G = (V, E)$ is connected (but not necessarily) a graph with a weighting function: $w: E \mapsto \mathbb{R}^+$. For simplicity, we can only say that the graph G w -weighted graph. For an arbitrary subset $E' \subseteq E$, with G' denotes subgraph of G induced by E' , and $w(E')$ denote the sum of weight on the vertices of E' , i.e. $w(E') = \sum_{e \in E'} w(e)$.

Let $G = (V, E)$ is an undirected graph, let $d \geq 2$ an integer, and let $w: E \mapsto \mathbb{R}^+$ weighting function. The problem limits the maximum degree of connected subgraph (MDBCS) consists in finding a subgraph $G' = (V', E')$, where $V' \subseteq V$ and $E' \subseteq E$, so that the subgraph G' is connected, and $\sum_{e \in E'} w(e)$ has a maximum value.

The constants in the ILP model are: E - denote arrays of sets, and values of functions w in them.

Variables for the ILP model are:

$$x_e = \begin{cases} 1, & e \in E' \\ 0, & e \notin E' \end{cases},$$

$$y_e = \begin{cases} 1, & e \in T' \\ 0, & e \notin T' \end{cases},$$

where $e \in E$, and T' is a spanning tree for the subgraph G' ,

$$z_i = \begin{cases} 1, & i \in V' \\ 0, & i \notin V' \end{cases}.$$

ILP formulation of the model for finding the maximum degree of a connected subgraph constraints is given below, part of the paper [4].

Determine

$$\max \sum_{e \in E} w_e x_e \quad (1)$$

with conditions:

$$\sum_{e \ni i} x_e \leq d, \quad \forall i \in V, \quad (2)$$

$$y_e \leq x_e, \quad \forall e \in E, \quad (3)$$

$$\sum_{e \in E} y_e = -1 + \sum_{i \in V} z_i, \quad (4)$$

$$x_e \leq z_{i_e}, \quad \forall e \in E, \quad (5)$$

$$x_e \leq z_{j_e}, \quad \forall e \in E, \quad (6)$$

$$\sum_{i_e, j_e \in S} y_e \leq |S| - 1, \quad \forall S \subseteq V, |S| \geq 3. \quad (7)$$

Graph G is not oriented. In the ILP model, the formula (5), indicates the starting node i_e of the vertice e , while in formula (6), j_e indicates an incoming (final) node of the vertice e .

The objective function, given by (1), maximizes the sum of the total weight.

Condition (2) ensures that the subgraph G' each node has the most d vertices leading from the G' .

The condition given by (3), ensures that the subgraph G' is a superset of T' and if T' as a spanning tree is connected, it follows that the subgraph G' also connected.

Conditions (4), (5) and (6), provided that the candidate for the spanning tree T' has as many nodes in the subgraph G' a vertices, minus 1. Finally, condition (7) guarantees that the candidate for the spanning tree T' has no cycle. Therefore, the conditions (4), (5), (6) and (7) together ensure that T' is the spanning tree for the subgraph G' .

The following theorem proves the correctness of the above ILP model.

Theorem 1. [4] The MDBCS problem can be solved if and only if the following conditions (2) - (7) holds or their equivalent set of conditions.

II. METAHEURISTIC GENETIC ALGORITHM FOR SOLVING THE MDBCS PROBLEM

Genetic algorithms (GA) are a family of algorithms, which use some of the genetic principles that are present in nature, in order to solve certain computational problems. These natural principles are: inheritance, crossover, mutation, survival of the best custom (survival of the fittest), migration and so on. These algorithms can be used for solving various classes of problems because they are fairly general nature. In this case, they are used in the optimization problem - finding the optimal parameters of a system ([5]).

In a narrow sense, the notion of a genetic algorithm applies only to the model introduced by John Holland in his book „Adaption in natural and artificial systems“, 1975. ([10]). Holland is considered the creator of this metaheuristic and basic settings of his earliest works are valid even today. In a broader sense, genetic algorithm is any algorithm that is based on a population and operators of selection, crossover and mutation, which are used to obtain new points in the search space.

Genetic algorithm is applied to the final set of individuals called the *population*. Each individual in the population is represented by a series of characters (genetic code) and corresponds to a solution in search space. Coding can be binary

or a transliteration of higher cardinality. Encoding solutions is an important step of genetic algorithm because, inadequate choice of code can lead to poor results regardless of the rest of the structure of the algorithm.

The diversity of genetic material is provided by generating the initial population randomly. It can be used and some heuristics to generate initial population, or part thereof, of course, if the heuristics perform relatively quickly and significantly if not reduce the diversity of genetic material.

Individuals assigned to each function adaptation (fitness function) that evaluates the quality of given individuals, as well as individual solutions in the search space. The task of genetic algorithms is to provide a constant, from generation to generation, improving the adaptability of the absolute population. This is realized by applying successive genetic operators selection, crossover and mutation, thus gaining a better solution given a particular problem.

Mechanism of selection favoring above average fitted individuals and their above-average custom parts (genes), which receive a higher chance of their own reproduction in the formation of a new generation. In this way, less fitted individuals and genes get less chances to play, and gradually dying out. Contribution to diversity of genetic material from your operator crosses that controls recombination genes of individuals. As a result of the crossing structure is obtained, although non-deterministic, the exchange of genetic material between individuals, with the possibility that well-adjusted individuals generate better individuals. The mechanism of crossing operators and relatively less fitted individuals, with some well-adapted genes, gets his chance to recombination of good genes produce well-adjusted individuals. However, using multiple selection and crossbreeding can result in loss of genetic material, ie, some regions of space results become available. The operator performs a random mutation of a particular gene, given the small probability p_{mut} , which can restore the lost genetic material in the population. This is the basic mechanism for preventing premature convergence of genetic algorithm to a local extreme, (see [3]).

Operators are applied until a stopping criterion is met, for example. reached the maximum number of generations, the same quality of solutions in a number of generations, finding the optimal solution, the best individual was repeated a maximum number of times, limited the time of executing the genetic algorithm, the termination by the user and so on.

As the most important aspects of a genetic algorithm, there are coding and fitness function, which is very important to be well adapted to the nature of a particular problem. It has been said that the usual binary encoding or over a large alphabet cardinality. The most convenient is that the relationship between the genetic codes and solutions to the problem is bijective mapping. Then it is possible that the application of genetic operators in a certain age to get called *incorrect specimen*, ie. individuals whose genetic code does not correspond to any solution. Overcoming this problem is possible in several ways. One possibility is to assign any such individuals as the fitness function value is zero, so that the operator applying for selection to eliminate these individuals. This approach has proven to be suitable only if the ratio of the

number of incorrect and correct individuals in the population is too large, which in practice often not the case. It is possible, however, incorrect inclusion of individuals in the population by the individuals assigned to each incorrect value penalty function. The aim is unfair to individuals and get a chance to participate in the crossing, but to be discriminated against on the correct individual. Care should be taken on how to balance the value of penalty function, because too small values can lead to a genetic algorithm that some of the incorrect code for a declaration of the solution, while, on the other hand, excessive punishment can cause loss of useful information from the incorrect individuals. There is another way to solve this problem - which is to improve specimen be unfair to make them correct or incorrect that each individual is replaced correctly.

Calculating the fitness function is possible in several ways. Some of these methods are direct download, linear scaling, interval scaling, sigma truncation, etc..

Since the selection is directly related to the fitness function, the basic way to implement this genetic operator is the simple roulette selection. This method uses a distribution where the probability of selection proportional to its adaptation to the individual. Individuals involved with the chances of roulette in accordance with them, pass or not pass the process of creating a new generation. The lack of a simple roulette selection is the possibility of premature convergence due to the gradual prevalence of highly adapted individuals in the population that do not correspond to the global optimum.

To avoid this problem can be used ranking selection based on genetic codes, according to their adaptability. Fitness function is equal to the individual a range of pre-specified number of ranks, and only depend on the position of individuals in the population. It can be used linearly, as well as other forms of ranking.

Another form of selection is the tournament selection. When tournament selection is randomly generated subsets of the N individuals (N is the pre-set number), then in each subset, the principle of the tournament, selects the best individual that participates in the creation of a new generation. Usually the problem is the choice of N so as to reduce the adverse effects of stochastic, so that better and more diverse genetic material passed to the next generation. In cases where the size is perfect tournament is not an integer, has proved successful fine-graded tournament selection (FGTS). A detailed description of these and other types of selection and its theoretical aspects can be found in [6]. Application of fine-graded tournament selection and comparison with other practices in the selection of operators are given in [7], [8], [9].

The process of exchange of genetic material between individuals of the parents, in order to form new offspring individuals, is performed by the crossover operator. The most common operators are one-point crossover, two-point crossover, multi-point and a uniform crossover, and can also be used for crossover mixing, reduced surrogate crossover, crossover with the mother, intermediate crossover, as well as the linear intersection.

Intersection operator, which is implemented in a simple genetic algorithm, the one-point crossover. In one-point crossover, so determined crossing position. All genes from the predetermined position, change position so that each parental pair created two offspring. In two-point crossover two positions are set and is the exchange of genetic material between the parents and two positions.

As for the uniform crossover, it should be noted that for each parental pair determines a binary string of length the same as the genetic parents. This range is called the mask. Sharing genes is performed only on those positions where the mask is 0, while in positions where there is one, the parents retain their genes.

Mutation operator is considered one of the most important and as such it can decisively influence the operation of genetic algorithm. If a genetic algorithm using binary encoding and the population of individuals not incorrect, it is usually implemented by a simple mutation operator that runs through the individual genes and the genetic code for each check whether or not mutated. The probability of mutation is pre-set p_{mut} small size, usually taken from the interval [0.001,0.01]. Simple mutation is sometimes possible to apply over a binary number - the mask, which is randomly generated for each individual, and carries information about the position in which the genetic code results in a change of genes.

When the gene encoding algorithm used whole or real numbers (floating point), it was necessary to develop other concepts of mutation, which was done. These are the replacement of genes randomly selected number (random replacement), add or subtract a small value (creep), multiply the number close to one (geometric creep) and so on. For both creep mutation operator required values are random and can have a uniform, exponential, Gaussian or binomial distribution (see [1], [2]).

In some cases it is useful to genes, depending on the position in the genetic code, have different levels of mutation. In this regard it is particularly important concept of frozen gene. Namely, if in a position of the genetic code in all or most of the population, the same gene, it is useful that the gene mutation has a higher level than the rest of the genetic code. This concept is used to restore lost diversity of genetic material, and these genes are called frozen.

Will the application of genetic algorithms have a success depends largely on the choice of replacement policy generation. Some of the most important policy of the replacement generation: generational genetic algorithm, genetic algorithm stationary and elitist strategy. Of course, it is possible to combine these principles.

Where the generational genetic algorithm, then apply to all individuals all the genetic operators, ie. there are no privileged individuals are going into the next generation, or individuals who go directly to the selection process.

On the contrary, stationary genetic algorithm favors the best individuals in the population so as to them shall not apply operator selection, but they go directly to the next stage, while the other applies the selection of individuals and they come to the remaining places.

Elitist strategy provides a direct passage into the next generation of one of the best individuals. These individuals do not apply to operators of selection, crossover and mutation. By applying genetic operators to the individuals of the population remaining seats are filled by the next generation.

This approach leaves room for another possible improvement of the genetic algorithm, which is caching. As an elite individuals pass from generation to generation unchanged, it and its value remains unchanged. Therefore, it would be useful to individuals elitinih value is remembered, rather than constantly calculated, saving the time required for their computation. This process is called caching, and more detail is in [12] and [11].

Namely, the calculated value of the objective function of individuals are stored in so-called. Hess-row table, which uses CRC codes that are assigned to individuals in population. If, during operation of the genetic algorithm, obtained through the same genetic code, then the objective function value is taken from a hash-table, through the CRC code.

Input for the algorithm for solving the MDBCS problem is undirected graph $G = (V, E)$, weight function (weighting function) $w: E \mapsto \mathbb{R}^+$ and integer $d \geq 2$. Each vertices of the graph encode the zero (0), if the condition (2) does not satisfy, and one (1), otherwise (we use that is, a binary encoding of individuals). If we find the node that has more than d vertices, then vertices of the node does not count, that is. set value 0.

In order to get connected components (or component connection), we apply a search in width. If the graph is only one component connection, then this is the end, that is, the graph is connected and calculate the sum of all vertices of the weight of the component. But if more than one connected components, then we take them in order, first and second, second and third, and so on, and look for the largest vertice of the weight (or cycle if it exists) that connects the two components. Then take all the vertices of the shortest path from the original graph, and the count only those with a degree $\leq d$. The process continues until they connect all components of relationship. Finally, we add all of this and get $\sum_{e \in E'} w(e)$.

Genetic operators, which are used here, are fine-graded tournament selection, a one-point crossover, a simple mutations with the frozen gene and caching techniques. Genetic algorithm is coded in the programming language C.

For checking the results of the implemented genetic algorithm based on mathematical models (1) - (7), we used the software package CPLEX. Genetic algorithm is tested on the test-examples reached the same values as the CPLEX program, and that the execution time was short in both cases, data on execution times not are presented.

As there are no standard instances of MDBCS problem, and the instance for KCT problem (k -cardinality tree problem), containing the nodes and weights that are appropriate for the considered problem, the necessary adjustments in accordance with the input of a genetic algorithm, are used in testing of the genetic algorithm for graphs with a large number of vertices and nodes.

Stopping criterion of the algorithm is the maximum number of generations 5000 or up to 2000 generations without improving the objective function value.

III. TEST EXAMPLES

Here are a few tables where the columns of labels, respectively, represent:

- instance name that contains the dimension of the input graph, an *Instance name*;
- cardinality of the set of nodes, n ;
- cardinality of the set of vertices, m ;
- integer $d \geq 2, d$;
- best solution obtained by genetic algorithm, GA_{best} ;
- average time t (in seconds) to calculate the best value;
- t_{tot} total time (in seconds) to complete a genetic algorithm;
- total number of generation, gen ;
- average value of using caching, *cache*.

The Tables 1 - 5 shows the results obtained by testing the genetic algorithm for instances with a large number of vertices and nodes, for different values of d .

IV. CONCLUSION

The problem of finding the maximum degree of limitation associated subgraph is very interesting, and then proposed a genetic algorithm can be the basis for further research and improvement. In addition, this class of combinatorial problems has been studied extensively and due to the application in the design of networks of networks and routing algorithms.

Some of the directions of further enlargement and improvement of the results could be:

- 1) the development of exact methods based on integer programming;
- 2) modification of metaheuristic described - genetic algorithm for solving similar problems on graphs;

3) obtaining new results from graph theory developed using the implementation.

REFERENCES

- [1] Beasley D., Bull D. R., Martin R. R., „ An Overview of Genetic Algorithms, Part1,“ Research Topics. University Computing; 1993, Vol. 15, No. 2, p. 58-69.
- [2] Beasley D., Bull D. R., Martin R. R., „ An Overview of Genetic Algorithms, Part2,“ Research Topics. University Computing; 1993, Vol. 15, No. 4, p. 170-181.
- [3] Bogdanović M., Rešavanje problema maksimalnog ograničenja stepena podgrafova u računarstvu, kao prilog teoriji grafova, Doktorska disertacija. Univerzitet u Beogradu, Matematički fakultet; 2010, PhD Thesis (in Serbian).
- [4] Bogdanović M., „ An ILP formulation for the maximum degree-bounded connected subgraph problem“, Computers & Mathematics with Applications; 2010, 59, No.9, p. 3029-3038.
- [5] Bogdanović M. „On some basic concepts of genetic algorithms as a meta-heuristic method for solving of optimization problems“, A Journal of Software Engineering and Applications; 2011, Vol. 4, No. 8, pp. 482-486, doi: 10.4236/jsea.2011.48055. Website: <http://www.scirp.org/journal/jsea>
- [6] Filipović V., Predlog poboljšanja operatora turnirske selekcije kod genetskih algoritama, Magistarski rad. Univerzitet u Beogradu, Matematički fakultet; 1998. MsThesis (in Serbian)
- [7] Filipović V., Kratica J., Tošić D., Ljubić I., „Fine Grained Tournament Selection for the Simple Plant Location Problem“, Proceedings on the 5th Online World Conference on Soft Computing Methods in Industrial Application – WSC5; 2000, p. 152-158.
- [8] Filipović V., Tošić D., Kratica J., “Experimental Results in Applying of Fine Grained Tournament Selection”, Proceedings of the 10th Congress of Yugoslav Mathematicians Belgrade, 21.-24.01.; 2001, p. 331-336.
- [9] Filipović, V., “Fine-Grained Tournament Selection Operator in Genetic Algorithms”, Computing and Informatics; 2003. 22(2), p.143-161.
- [10] Holland J. H., Adaptation in Natural and Artificial Systems. The University of Michigan Press, Ann Arbor; 1975.
- [11] Kratica J., „ Improvement of Simple Genetic Algorithm for Solving the Uncapacitated Warehouse Location Problem“, Advances in Soft Computing – Engineering Design and Manufacturing, R. Roy, T. Furuhashi and P. K. Chawdhry (Eds), Springer-Verlang London Limited; 1999, p. 390-402.
- [12] Kratica J., Paralelizacija genetskih algoritama za rešavanje nekih NP-kompletnih problema, Doktorska disertacija. Matematički fakultet, Beograd; 2000, PhD Thesis (in Serbian).

TABLE 1.

<i>Instance name</i>	<i>n</i>	<i>m</i>	<i>d</i>	GA_{best}	$t[s]$	$t_{tot}[s]$	<i>gen</i>	<i>cache [%]</i>
Proba10-13	10	13	2	60	0.0001	0.50	2003	97.94716
Proba10-13	10	13	3	71	0.0001	0.53	2004	98.11958
Proba10-13	10	13	4	74	0.0001	0.48	2005	98.19920
Proba10-13	10	13	5	75	0.0001	0.50	2004	98.18934
Proba10-13	10	13	6	75	0.0001	0.48	2004	98.19934

TABLE 2.

<i>Instance name</i>	<i>n</i>	<i>m</i>	<i>d</i>	<i>GA_{best}</i>	<i>t[s]</i>	<i>t_{tot}[s]</i>	<i>gen</i>	<i>cache [%]</i>
w40-40	40	40	2	576	0.02	0.53	2042	75.21076
w40-40	40	40	3	768	0.09	0.64	2298	80.43894
w40-40	40	40	4	906	0.01	0.55	2040	82.20656
w40-40	40	40	5	1013	0.05	0.59	2114	82.38545
w40-40	40	40	6	1069	0.01	0.55	2048	82.72843
w40-40	40	40	7	1111	0.01	0.53	2037	85.22745
w40-40	40	40	8	1120	0.0001	0.52	2026	85.13061
w40-40	40	40	9	1120	0.0001	0.52	2026	85.13061
h44-44	44	44	2	85	0.08	0.67	2032	57.63736
h44-44	44	44	3	203	0.02	0.53	2050	79.11934
h44-44	44	44	4	292	0.05	0.56	2137	77.93271
h44-44	44	44	5	343	0.03	0.50	2112	81.25106
h44-44	44	44	6	371	0.01	0.55	2042	83.19511
h44-44	44	44	7	377	0.01	0.53	2038	85.13866
h44-44	44	44	8	377	0.01	0.53	2038	85.13866
w65-65	65	65	2	283	0.25	1.11	2456	43.06466
w65-65	65	65	3	787	0.42	1.22	3004	52.15896
w65-65	65	65	4	1200	0.47	1.28	3101	55.82088
w65-65	65	65	5	1488	0.05	0.75	2093	68.62977
w65-65	65	65	6	1644	0.34	1.03	2913	71.04870
w65-65	65	65	7	1744	0.03	0.72	2060	72.59913
w65-65	65	65	8	1792	0.03	0.70	2058	76.46288
w65-65	65	65	9	1801	0.05	0.72	2077	75.60385
w65-65	65	65	10	1801	0.05	0.72	2077	75.60385

TABLE 3.

<i>Instance name</i>	<i>n</i>	<i>m</i>	<i>d</i>	<i>GA_{best}</i>	<i>t[s]</i>	<i>t_{tot}[s]</i>	<i>gen</i>	<i>cache [%]</i>
h118-118	118	118	2	127	0.01	1.33	2007	33.67463
h118-118	118	118	3	601	0.69	2.00	2999	39.21786
h118-118	118	118	5	1235	2.59	3.34	5000	44.09634
h118-118	118	118	6	1663	3.06	3.19	5000	49.11733
h118-118	118	118	7	1958	0.41	1.48	2648	59.18974
h118-118	118	118	8	2058	0.11	1.17	2143	61.19571
h118-118	118	118	9	2091	0.09	1.17	2117	61.18774
h118-118	118	118	10	2091	0.09	1.17	2117	61.18774
w130-130	130	130	2	295	0.02	1.38	2017	34.75842
w130-130	130	130	3	1030	1.30	2.70	3788	34.68636
w130-130	130	130	4	1657	0.31	1.75	2406	36.35866
w130-130	130	130	5	2084	0.88	2.28	3195	39.45466
w130-130	130	130	6	2323	0.64	2.08	2877	38.50201
w130-130	130	130	7	2481	0.13	1.50	2152	42.08631
w130-130	130	130	8	2582	0.36	1.67	2488	46.23685
w130-130	130	130	9	2591	0.11	1.41	2117	46.92453
w130-130	130	130	10	2591	0.11	1.41	2117	46.92453
h183-183	183	183	2	169	0.01	1.83	2019	32.26805
h183-183	183	183	3	1000	2.47	4.36	4525	32.48498
h183-183	183	183	4	2056	2.20	4.16	4167	35.62782
h183-183	183	183	5	3007	2.77	4.66	4780	40.11583
h183-183	183	183	6	3776	0.83	2.48	2892	50.47738
h183-183	183	183	7	4147	2.09	3.75	4392	51.32014
h183-183	183	183	8	4343	1.59	3.23	3829	51.77192
h183-183	183	183	9	4425	0.27	1.92	2232	51.54989
h183-183	183	183	10	4457	0.38	2.00	2367	51.83713
h183-183	183	183	11	4457	0.39	2.02	2367	51.83713

TABLE 4.

Instance name	<i>n</i>	<i>m</i>	<i>d</i>	<i>GA_{best}</i>	<i>t</i> [s]	<i>t_{tot}</i> [s]	<i>gen</i>	<i>cache</i> [%]
h212-212	212	212	2	156	0.22	2.30	2179	32.14757
h212-212	212	212	3	1203	1.52	3.75	3230	33.58119
h212-212	212	212	4	2624	4.09	5.84	5000	33.34679
h212-212	212	212	5	3959	4.95	5.69	5000	37.96402
h212-212	212	212	6	4902	2.72	4.69	4645	47.71343
h212-212	212	212	7	5376	2.55	4.50	4502	49.54007
h212-212	212	212	8	5571	1.33	3.25	3279	49.15052
h212-212	212	212	9	5654	2.20	4.09	4225	50.38174
h212-212	212	212	10	5686	0.38	2.27	2291	48.96513
h212-212	212	212	11	5686	0.36	2.25	2291	48.96513
h44-48	44	48	2	177	0.27	0.86	2826	61.00530
h44-48	44	48	3	335	0.48	1.22	3799	78.49343
h44-48	44	48	4	405	0.05	0.61	2110	72.81590
h44-48	44	48	5	452	0.19	0.77	2679	73.40791
h44-48	44	48	6	477	0.05	0.59	2147	77.68186
h44-48	44	48	7	484	0.02	0.55	2045	78.76856
h44-48	44	48	8	489	0.01	0.55	2032	82.67125
h44-48	44	48	9	489	0.01	0.55	2032	82.67125
w70-76	70	76	2	625	0.11	0.97	2225	44.62029
w70-76	70	76	3	1307	0.31	1.20	2664	50.06899
w70-76	70	76	4	1732	0.09	0.89	2157	62.85833
w70-76	70	76	5	1991	0.06	0.81	2096	69.20438
w70-76	70	76	6	2109	0.08	0.80	2099	69.92674
w70-76	70	76	7	2159	0.66	1.42	3702	69.68583
w70-76	70	76	8	2203	0.03	0.80	2068	70.37373
w70-76	70	76	9	2224	0.03	0.75	2064	72.95791
w70-76	70	76	10	2224	0.03	0.77	2064	72.95791

TABLE 5.

Instance name	<i>n</i>	<i>m</i>	<i>d</i>	<i>GA_{best}</i>	<i>t</i> [s]	<i>t_{tot}</i> [s]	<i>gen</i>	<i>cache</i> [%]
proba12-17	12	17	2	118	0.02	0.53	2010	95.20616
proba12-17	12	17	3	128	0.0001	0.55	2011	96.62761
proba12-17	12	17	4	128	0.0001	0.52	2011	96.62761
g15-4-01	15	20	2	326	0.0001	0.52	2011	94.55214
g15-4-01	15	20	3	335	0.0001	0.53	2011	95.36941
g15-4-01	15	20	4	335	0.0001	0.52	2011	95.36941
liter34-39	34	39	2	282	0.0001	0.56	2018	67.02524
liter34-39	34	39	3	349	0.01	0.53	2028	84.96898
liter34-39	34	39	4	369	0.09	0.66	2363	81.58072
liter34-39	34	39	5	371	0.01	0.53	2033	84.18566
liter34-39	34	39	6	371	0.01	0.53	2033	84.18566
h69-69	69	69	2	382	0.13	0.95	2248	49.79209
h69-69	69	69	3	572	0.28	1.03	2680	62.79687
h69-69	69	69	4	690	0.33	1.03	2781	69.99425
h69-69	69	69	5	752	0.16	0.84	2351	71.30161
h69-69	69	69	6	771	0.03	0.72	2064	74.50798
h69-69	69	69	7	771	0.03	0.72	2064	74.50798
h148-152	148	152	2	2107	0.77	2.45	2829	38.28884
h148-152	148	152	3	3470	2.16	3.66	4685	48.01664
h148-152	148	152	4	3811	0.56	1.97	2688	53.01078
h148-152	148	152	5	3931	0.17	1.61	2178	51.45988
h148-152	148	152	6	3941	0.19	1.56	2191	54.73291
h148-152	148	152	7	3941	0.19	1.56	2191	54.73291

Optimized Min-Sum Decoding Algorithm for Low Density Parity Check Codes

Mohammad Rakibul Islam, Dewan Siam Shafiullah, Muhammad Mostafa Amir Faisal, Imran Rahman
Dept. of Electrical & Electronic Engineering, Islamic University of Technology
Boardbazar, Gazipur-1704, Dhaka, Bangladesh

Abstract — Low Density Parity Check (LDPC) code approaches Shannon-limit performance for binary field and long code lengths. However, performance of binary LDPC code is degraded when the code word length is small. An optimized min-sum algorithm for LDPC code is proposed in this paper. In this algorithm unlike other decoding methods, an optimization factor has been introduced in both check node and bit node of the Min-sum algorithm. The optimization factor is obtained before decoding program, and the same factor is multiplied twice in one cycle. So the increased complexity is fairly low. Simulation results show that the proposed Optimized Min-Sum decoding algorithm performs very close to the Sum-Product decoding while preserving the main features of the Min-Sum decoding, that is low complexity and independence with respect to noise variance estimation errors.

Keywords — LDPC codes; Min-sum algorithm; Normalized min-sum algorithm; Optimization factor.

I. INTRODUCTION

Among the error correction codes, Low Density Parity Check (LDPC) is one of the most efficient techniques. It was first introduced by Robert Gallager in 1962 in his PhD. Dissertation [1]. It is the extreme sparseness of the parity check matrix for LDPC codes that make the decoding particularly attractive. LDPC codes have recently received a lot of attention because they can achieve a remarkable performance near Shannon limit over the binary symmetric channel (BSC) as well as the additive white Gaussian noise (AWGN) channel [2]. The decoding of an LDPC code allows a high degree of parallelism, which makes it very suited for high data rate applications such as wide-band wireless multimedia communications and magnetic storage systems [3], [4]. The low-density nature of the parity check matrix thus contributes both to good distance properties and the relatively low complexity of the decoding algorithm [5]. Well-designed irregular LDPC codes demonstrate better performance than regular ones [6].

Among a variety of decoding algorithms, the well-known Sum Product (SP) algorithm [7] achieves a good decoding performance but requires a large hardware complexity. There are alternative methods such as several kinds of Min-Sum (MS) algorithms which can significantly reduce the hardware complexity of SP at the cost of acceptable performance degradation where complex computations at the check nodes are approximated by using simple comparison and summation operations. Recently, the modified MS algorithms using

correction factors have been preferred for many practical applications since they offer comparable decoding performance compared to that of SP [7] for regular LDPC codes [8], [9]. Also, for irregular LDPC codes, the improved normalized or offset MS algorithms exhibit small performance degradations [10], [11]. Specifically, the offset MS algorithm has been implemented for several practical applications due to its better performance and simple computations.

The main decoding algorithms of LDPC codes include soft-decision such as Sum Product (SP) algorithm [7] and hard-decision such as Bit flipping. In iterative decoding, a critical tradeoff between "complexity" and "performance" is required. Based on these two issues, LDPC codes may be classified as optimal, sub-optimal or quasi-optimal. The optimal iterative decoding is performed by the Sum-Product algorithm [7] at the price of an increased complexity, computation instability, and dependence on thermal noise estimation errors. The Min-Sum algorithm [12] performs a suboptimal iterative decoding, less complex than the Sum-Product decoding. The sub-optimality of the Min-Sum decoding comes from the overestimation of check-node messages, which leads to performance loss with respect to the Sum-Product decoding. Several correction methods were proposed [13-15] in the literatures in order to recover the performance loss of the Min-Sum decoding with respect to the Sum-Product decoding which are called quasi-optimal algorithms. An example is Normalized min-sum algorithm proposed by Chen and Fossorier [16]. In this paper, we propose an optimized min-sum algorithm which has better performance not only from min-sum algorithm but also from normalized min-sum algorithm.

The rest of the paper is organized as follows. In section II, different LDPC decoding algorithms are discussed, and section III explains our proposed Optimized Min-sum algorithm. Section IV discusses the simulation results, and finally section V concludes the paper.

II. LDPC DECODING ALGORITHMS

Decoding of LDPC codes can be two types: hard decision decoding and soft decision decoding.

1) Hard Decision Decoding

For each bit c_n , compute the checks for those checks that are influenced by c_n . If the number of nonzero checks exceeds some threshold (s_n , the majority of the checks are nonzero), then the bit is determined to be incorrect. The erroneous bit is flipped, and correction continues. This simple scheme is

capable of correcting more than one error. Suppose that c_n is in error and that other bits influencing its checks are also in error. Arrange the Tanner graph with c_n as a root considering no cycle in the graph. In Fig. 1, suppose the bits in the shaded boxes are in error. The bits that connect to the checks connected to the root node are said to be in tier 1. The bits that connect to the checks from the first tier are said to be in tier 2. Then, decode by proceeding from the “leaves” of the tree (the top of the figure). By the time decoding on c_n is reached, other erroneous bits may have been corrected. Thus, bits and checks which are not directly connected to c_n can still influence c_n .

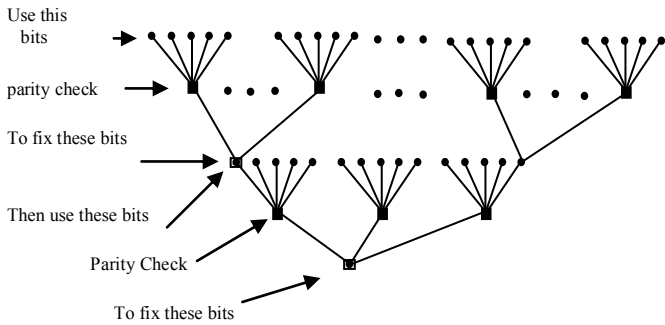


Figure 1. A parity check tree associated with the Tanner graph [18]

2) Soft Decision Decoding

In the Soft decision decoding, rather than flipping bits (a hard operation), we propagate probabilities through the Tanner graph, thereby accumulating evidence that the checks provide about the bits. The optimal (minimum probability of decoding error) decoder seeks a codeword \hat{c} which maximizes $P(\hat{c} | r, A\hat{c} = 0)$. So, it seeks the most probable vector which satisfies the parity checks, given set of received data $r = [r_1, r_2, \dots, r_N]$.

However, the decoding complexity for the true optimum decoding of an unstructured (i.e., random) code is exponential in K , requiring an exhaustive search over all 2^k codewords. Instead, the decoder attempts to find a codeword having bits c_n which maximize $P(c_n | r, \text{all checks involving bit } c_n \text{ are satisfied})$, it is the posterior probability for a single bit given that only the checks on that bit are satisfied. As it turns out, even this easier, more computationally localized, task cannot be exactly accomplished due to approximations the practical algorithm must make. However, the decoding algorithm has excellent performance and the complexity of the decoding is linear in the code length.

LDPC decoding is based on the parity check matrix which can also be represented using a bipartite graph. Columns in the parity check matrix represent variable nodes and rows in the matrix represent check nodes. Each variable node corresponds to one bit of the codeword and each check node corresponds to one parity check equation. Edges in the graph connect variable nodes to check nodes and represent the nonzero entries in H matrix. The term “low density” conveys the fact that the fraction of nonzero entries in H is small, in particular it is linear in the block length n . Parity check matrix can be of regular and irregular types. In this paper, we use the regular codes. For regular codes, the corresponding H matrix has d_c ones in each row and d_v ones in each column. It means that every codeword bit participates in exactly d_c parity check equations and that

every such check equation involves exactly d_v codeword bits. Low density parity check codes have been constructed mostly using regular random bipartite graphs, here is an example of a regular parity check matrix with $d_c = 3$ and $d_v = 3$.

$$H = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

A graph associated with a parity check matrix A is called the Tanner graph and it contains two sets of nodes. The first set consists of N nodes which represent the N bits of a codeword; nodes in this set are called “bit” nodes. The second set consists of M nodes, called “check” nodes representing the parity constraints. The graph has an edge between the n th bit node and the m -th check node if and only if n th bit is involved in the m th check, that is, if $A_{mn} = 1$. Thus, the Tanner graph is a graphical depiction of the parity check matrix. The bipartite graph corresponding to this parity check matrix is shown in Fig. 2.

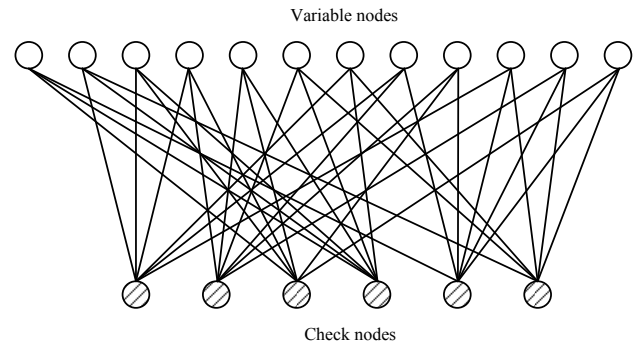


Figure 2. Bipartite graph corresponding to a regular parity check matrix

Let C be a regular LDPC code of length N and dimension K whose parity-check matrix A with $M = N - K$ rows and N columns contains exactly d_v 1's in each column (column weight) and exactly d_c 1's in each row (row weight)

A_{mn} is the value of the m th row and n th column in A . The set of bits that participate in check is denoted: $N_m = \{n: A_{mn} = 1\}$. The set of checks that participate in bits $M_n = \{m: A_{mn} = 1\}$.

Assume codeword, $c = [c_1, c_2, c_3, \dots, c_N]^T$. Before transmission, it is mapped to a signal constellation to obtain the vector, $t = [t_1, t_2, t_3, \dots, t_N]^T$,

where

$$t_n = 2 * c_n - 1,$$

which is transmitted through an AWGN channel with variance

$$\sigma^2 = N_0/2,$$

$$r = [r_1, r_2, r_3, \dots, r_N]^T$$

where

$$r_n = t_n + v_n.$$

Here, v_n is the Additive White Gaussian Noise (AWGN) with zero mean. Let hard decision vector,

$$z = [z_1, z_2, z_3 \dots \dots, z_N]^T \text{ Be } z_n = \text{sgn}(r_n)$$

$$\text{Where } \text{sgn}(r_n) = \begin{cases} 1 & r_n > 0 \\ 0 & \text{otherwise} \end{cases}$$

The following notations concern bipartite graphs and message-passing algorithms running on these graphs and will be used throughout the paper.

L_n : A priori information of bit node, n

\bar{L}_n : A posteriori information of bit node, n

$E_{m,n}$: The check to bit message from m to

$F_{n,m}$: The bit to check message from n to m

A. *Sum Product Algorithm:*

The Sum Product Algorithm [17] can be summarized in the following four steps.

Step 1: Initialization

A priori information, $L_n = -r_n$

Bit to check message initialization, $F_{n,m} = L_n$

Step 2: Horizontal Step

Check node Processing:

$$E_{m,n} = \log \frac{1 + \prod_{n' \in N(m) \setminus n} \tanh\left(\frac{F_{n',m}}{2}\right)}{1 - \prod_{n' \in N(m) \setminus n} \tanh\left(\frac{F_{n',m}}{2}\right)}$$

Step 3: Vertical Step

A posteriori information:

$$\bar{L}_n = L_n + \sum_{m \in M(n)} E_{m,n}$$

Bit node Processing:

$$F_{n,m} = \bar{L}_n + \sum_{m' \in M(n) \setminus m} E_{m',n}$$

Step 4: Decoding Attempt

$\bar{L}_n > 0, \bar{c}_n = 0$, else $\bar{c}_n = 1$

If $A\bar{c}_n = 0$ then the algorithm stops and \bar{c}_n is considered as a valid decoding result.

Otherwise, it goes to next iteration until the number of iteration reaches its maximum limit.

B. Log Likelihood Decoding Algorithm for Binary LDPC codes

Step 1: Initialization:

Set $\eta_{m,n}^{[0]} = 0$ for all (m, n) with $A(m, n) = 1$.

Set $\lambda_{mn} = L_c r_n$

Set the loop counter $l = 1$.

Step 2: Check node update:

For each (m, n) with $A(m, n) = 1$, Compute

$$\eta_{m,n} = 2 \tanh^{-1} \left(\prod_{j \in N_{m,n}} \tanh \left(\frac{\lambda_{mj}}{2} \right) \right)$$

Step 3: Bit node update: For each (m, n) with $A(m, n) = 1$, Compute

$$\lambda_{mn} = L_c r_n + \sum_{m \in M_{n,m}} \eta_{m,n}$$

Log pseudo posterior probabilities:

For $n = 1, 2 \dots, N$ Compute,

$$\lambda_n = L_c r_n + \sum_{m \in M_n} \eta_{m,n}$$

Step 4: Make a tentative decision: Set $\hat{c}_n = 1$ if $\lambda_n > 0$, else set, $\hat{c}_n = 0$

If $A\hat{c}_n = 0$ then stop, otherwise, if the number of iteration < maximum number of iteration, loop to check node update. Otherwise, declare decoding failure and stop.

C. Min Sum Decoding

The sum-product algorithm can be modified to reduce the implementation complexity of the decoder.

This can be done by altering the Horizontal step:

$$E_{m,n} = \log \frac{1 + \prod_{n' \in N(m) \setminus n} \tanh\left(\frac{F_{n',m}}{2}\right)}{1 - \prod_{n' \in N(m) \setminus n} \tanh\left(\frac{F_{n',m}}{2}\right)} \quad (1)$$

using the relationship:

$$2 \tanh^{-1} p = \log \frac{1+p}{1-p}$$

Equation (1) can be rewritten as,

$$E_{m,n} = 2 \tanh^{-1} \prod_{n' \in N(m) \setminus n} \tanh\left(\frac{F_{n',m}}{2}\right) \quad (2)$$

Equation (2) can be further modified as,

$$E_{m,n} = 2 \tanh^{-1} \prod_{n' \in N(m) \setminus n} \text{sgn}(F_{n',m}) \prod_{n' \in N(m) \setminus n} \tanh\left(\frac{|F_{n',m}|}{2}\right)$$

$$= \prod_{n' \in N(m) \setminus n} \text{sgn}(F_{n',m}) 2 \tanh^{-1} \prod_{n' \in N(m) \setminus n} \tanh\left(\frac{|F_{n',m}|}{2}\right) \quad (3)$$

The Min-sum algorithm simplifies the calculation of (3) even further by recognizing that the term corresponding to the smallest $F_{n,m}$ dominates the product term and so the product can be approximated by a minimum:

$$E_{m,n} = \prod_{n' \in N(m) \setminus n} \text{sgn}(F_{n',m}) \min_{n' \in N(m) \setminus n} |F_{n',m}| \quad (4)$$

D. Normalized Min Sum Decoding

Normalized Min sum algorithm [16] further modifies the min sum algorithm by multiplying a normalizing factor (say ν) where $0 < \nu \leq 1$ in the horizontal step to achieve a better error performance closer to sum product algorithm.

$$E_{m,n} = \nu \prod_{n' \in N(m) \setminus n} \text{sgn}(F_{n',m}) \min_{n' \in N(m) \setminus n} |F_{n',m}| \quad (5)$$

A flow chart of Normalized Min Sum algorithm is given below:

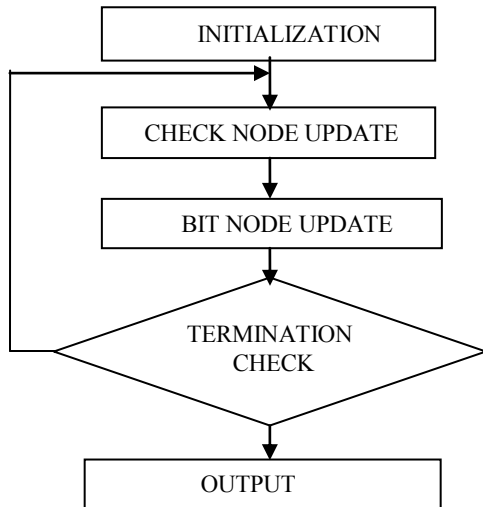


Figure 3. Flow chart of Normalized Min Sum Algorithm.

III. PROPOSED OPTIMIZED MIN-SUM ALGORITHM

1) Motivation

The foundation of our work is based on the improvement in error performance of normalized min sum algorithm [16]. From descriptions in previous sections, we have seen that the sum product decoding [7] has been reduced to different forms to reduce the complexity and through some compromise in performance. Min sum decoding [12] algorithm is one of them. Different works have been done on min sum decoding to improve its performance to get closer to sum product algorithm performance like normalized min sum decoding algorithm[16], adaptive min sum decoding algorithm[14], self-corrected min sum decoding algorithm[15] etc. In these papers, they proposed different factors which modifies and improves the error performance in different ways. In the normalized min-sum algorithm, a normalizing factor was proposed to be multiplied in check node. But, the error performance using normalized min-sum algorithm can be further modified to get closer to the error performance of sum product algorithm.

2) Optimization Factor, α

The value of optimization factor α varies for different Signal to Noise Ratio (SNR). For a particular SNR, we took the value of α that causes the minimum Bit Error Rate (BER).

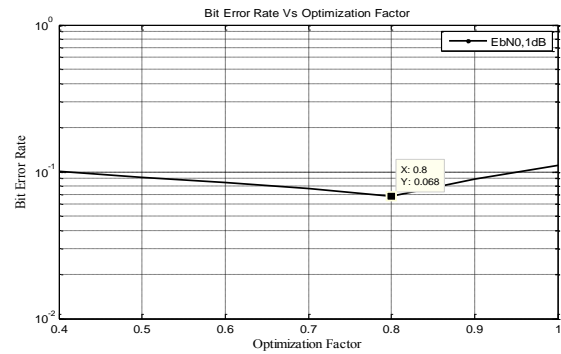


Figure 4. The impact of the optimization factor in the Optimized Min-Sum algorithm on the BER for the (2000, 1000) LDPC codes

Fig. 4 shows the variation of BER with respect to optimization factor, α for 1dB Signal to Noise Ratio. Here $\alpha=0.8$ is selected for which the BER is minimum. This same procedure is followed to calculate α for different SNRs.

3) Proposed Algorithm

In line with our motivation and the previously explained normalized min-sum algorithm, we propose the optimized min-sum algorithm. The main feature of our proposed algorithm is the use of the optimization factor. Multiplication of α both in check node and bit node update is the basic difference between optimized min-sum algorithm and Normalized Min-sum algorithm. In the Normalized Min-sum algorithm, normalizing factor was used for check node update only [16]. Also in 2 Dimensional Normalized Min Sum algorithm [19], two different factors for check and bit node updates are used and multiplied in 3 different places, check node processing, A posteriori information and bit node processing. The advantage of the proposed algorithm is that only the optimization factor is used for both bit node and check node updates. Also, the Optimization factor is not multiplied in a posteriori information which reduces complexity of the algorithm. The proposed algorithm is explained in Fig. 5 where a flow chart is shown.

First we initialize the bit to check message. Then we update the check message in the horizontal step. In this step, we multiply the Optimization factor α with the check message. After that, we proceed to the vertical step. In this step, we update the posteriori information with the help of check message and then we update the bit node. Here, we multiply the Optimization factor α with the check message. The last step is the decision making process. If the decoded codeword is correct, we stop there and take it as the output or otherwise repeat the whole decoding process until the iteration number reaches its maximum limit.

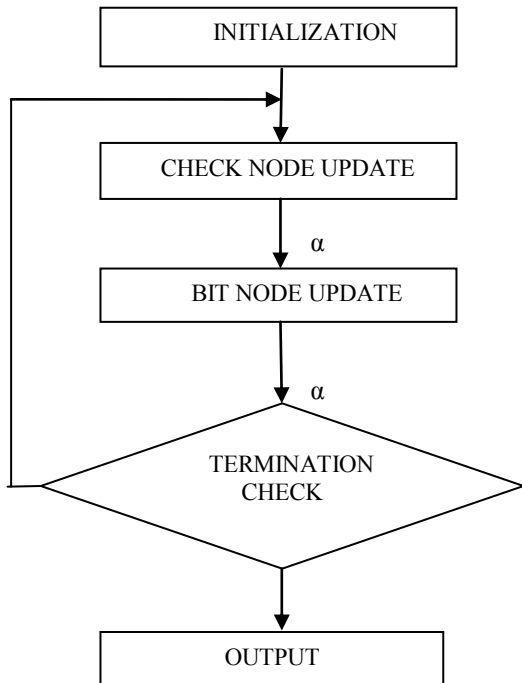


Figure 5. Flow chart of Optimized Min-Sum Algorithm

The detailed version of the algorithm is shown in the following steps. α is the optimization factor whose range is $0 < \alpha \leq 1$.

Step 1: Initialization

A priori information, $L_n = -r_n$

Bit to check message initialization, $F_{n,m} = L_n$

Step 2: Horizontal step

Check node processing:

$$E_{m,n} = \alpha \prod_{n' \in N(m) \setminus n} \text{sgn}(F_{n',m}) \min_{n' \in N(m) \setminus n} |F_{n',m}| \quad (6)$$

Step 3: Vertical step

A posteriori information:

$$\bar{L}_n = L_n + \sum_{m \in M(n)} E_{m,n} \quad (7)$$

Bit node processing:

$$F_{n,m} = \bar{L}_n - \alpha E_{m,n} \quad (8)$$

Step 4: Decoding Attempt

If $\bar{L}_n > 0$, $\bar{c}_n = 0$,

else $\bar{c}_n = 1$

If $\bar{A}c_n = 0$

Then the algorithm stops and \bar{c}_n is considered as a valid decoding result. Otherwise, it goes to next iteration until the number of iteration reaches its maximum limit.

IV. SIMULATION RESULTS

A. Error Performance Analysis

In total, we observed 4 simulations. The first one is regular (1944, 972) LDPC codes for IEEE 802.16e with code rate 1/2, row weight 7 and column weight 11. The codes are transmitted on AWGN channel after BPSK modulation. We set the maximum number of iteration to 50. The comparison among Sum Product (SP) algorithm [7], Min-sum (MS) algorithm [12], Normalized Min-sum (NMS) algorithm [16], 2 Dimensional Normalized Min Sum (2D NMS) algorithm and proposed Optimized Min-Sum (OMS) algorithm are shown in the Fig. 6. Simulation results show that the Optimized Min-sum algorithm obtains much better performance than Min sum algorithm, comparatively better performance than Normalized min sum algorithm, 2 Dimensional Normalized Min Sum (2D NMS) algorithm and closer to that of Sum Product algorithm. Fig. 6 shows that for the BER value, 10^{-3} , our algorithm can achieve 0.05dB decoding gain over 2Dimensional Normalized Min Sum algorithm and 0.1dB gain over Normalized Min Sum algorithm.

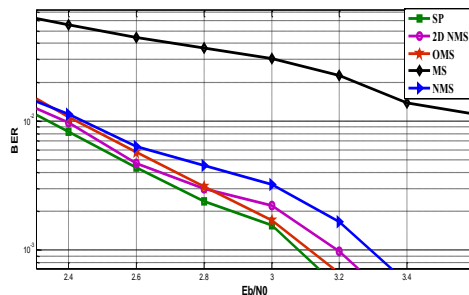


Figure 6. Bit Error Rate of LDPC codes (1944, 972) for SP, MS, NMS, and OMS

The second one is regular (1944, 1296) LDPC codes with code rate 2/3, row weight 11 and column weight 8 are used. The codes are also transmitted on AWGN channel after BPSK modulation and we set the maximum number of iteration to 50. The comparison among Sum Product (SP) [7], Min-sum (MS) [12], Normalized Min-sum (NMS) [16] and proposed Optimized Min-Sum (OMS) algorithms are shown in the Fig. 7.

Simulation results in Fig. 7 show that for the BER value, 10^{-2} , our algorithm can achieve around 0.3dB decoding gain over Normalized Min sum algorithm which depicts that Optimized Min-sum algorithm significantly better than Normalized Min sum algorithm in error performance.

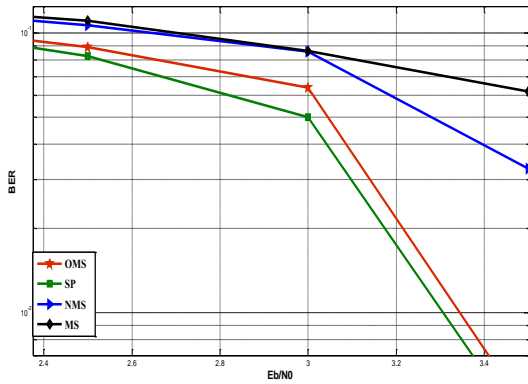


Figure 7. Bit Error Rate of LDPC codes (1944, 1296) for SP, MS, NMS, and OMS

The third one is regular (1944, 1458) LDPC codes with code rate 3/4, row weight 14 and column weight 6 are used. The codes are transmitted again on AWGN channel after BPSK modulation and we set the maximum number of iteration to 50. The comparison among Sum Product (SP) algorithm [7], Min-sum (MS) algorithm [12], Normalized Min-sum (NMS) algorithm [16] and proposed Optimized Min-Sum (OMS) algorithms are shown in the Fig. 8. Simulation results show that the Optimized Min-sum obtains much better performance than Min sum algorithm, comparatively better performance than Normalized Min-sum algorithm and closer to that of Sum Product. Fig. 8 shows that for the BER value, 10^{-2} , our algorithm can achieve 0.2dB decoding gain over Normalized Min-Sum algorithm.

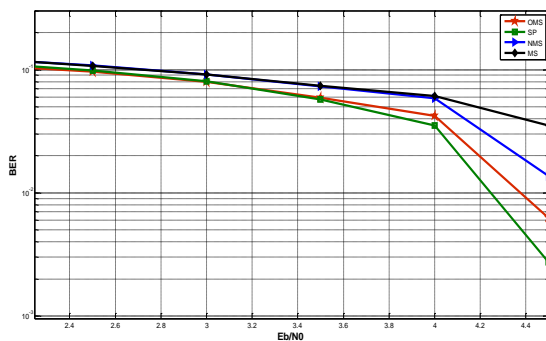


Figure 8. Bit Error Rate of LDPC codes (1944, 1458) for SP, MS, NMS and OMS

From the figures, it is clear that Optimized Min-Sum algorithm consistently shows better performance from Normalized Min-Sum algorithm and Min-Sum algorithm for different rates. We can also notice that for 1/2 rate code, Optimized Min-Sum algorithm has only -0.03dB gain over Sum Product algorithm and for 2/3 and 3/4 code rates around -0.05dB gain and -1dB gain respectively, So, if we take in account the reduction of complexity, it can be said that Optimized Min-Sum algorithm is almost comparable to Sum Product decoding algorithm.

The earlier simulations were run for AWGN channels. Practically, fading exists in channels. There are various types

of fading channel models i.e. Rayleigh, Weibul, Log Normal etc. So to get a more practical view of Optimized Min-Sum Algorithm, error performance of the algorithm in AWGN, Rayleigh, Weibul and Log Normal channels are compared.

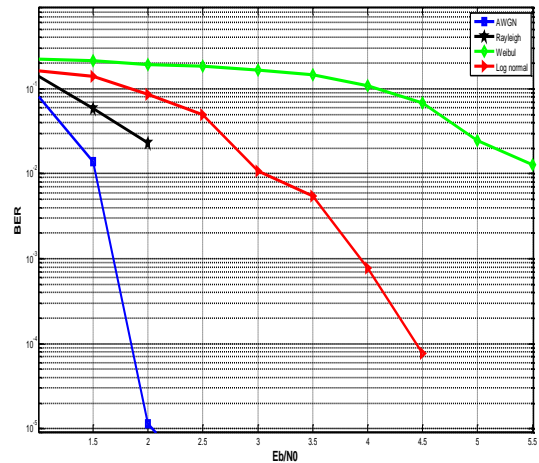


Figure 9. Comparison of OMS in AWGN, Rayleigh, Weibul and Log normal channels.

Fig 9 shows that the error performance varies with variation in fading. Error performance is best when fading is ignored in AWGN channel. In case of other channel models with fading, error performance degrades according to the degree of fading.

B. Complexity Analysis

TABLE I. COMPLEXITY CALCULATION

Name of the Algorithms	Calculations	
	Addition	Multiplication
Sum Product Decoding	150	2150
Min-Sum Algorithm	150	1100
Normalized Min-Sum Algorithm	60	1250
Optimized Min-Sum Algorithm	60	1400
2D Normalized Min-Sum	60	1650

Optimized Min-Sum algorithm is a quasi optimal decoding algorithm which improves error performance from Min-Sum decoding algorithm through slight increase in complexity. A (6, 3) regular LDPC code of code rate 1/2 was used to compare the complexity among different algorithms. Optimal Sum Product (SP) Algorithm has highest complexity and the Sub Optimal Min Sum (MS) algorithm has the lowest complexity. For quasi optimal codes, a good tradeoff between 'complexity' and 'Error Performance' is required, Normalized Min Sum (NMS) algorithm improves error performance from Min-Sum (MS) algorithm but, complexity increases as the table shows an increase in multiplication. For the proposed Optimized Min-Sum Algorithm table shows slight increase in multiplication

because of using the Optimization factor α for two updates but the tradeoff between error performance and complexity is an attractive as can be seen from the previous section. 2D Normalized Min-Sum algorithm has further increase in complexity than Optimized Min-Sum algorithm due to additional multiplication in the a posteriori information as can be seen from the table.

V. CONCLUSION

Through the introduction of Optimization factor we have obtained a better tradeoff between 'performance' and 'complexity'. We achieved much better performance than Min Sum algorithm and Normalized Min Sum algorithm in exchange of a slight increase in complexity. Using the same factor for both the nodes has reduced the complexity of calculating two different factors.

The optimization factor is determined before the decoding process which causes no additional complexity in the decoding algorithm. Thus, the proposed algorithm is a very competitive decoding technique for regular LDPC codes. Further analysis can be done for irregular LDPC codes and hardware implementation is also possible.

REFERENCES

- [1] R. G. Gallager, "Low-density parity check codes," IRE Trans. on Information Theory, vol. IT-8, pp.21-28, Jan. 1962.
- [2] O. J. C. Mackay, "Good error correcting codes based on very sparse matrices," IEEE Trans. on Inform. Theory, vol. 45, pp.399-431, Mar. 1999.
- [3] M. Rovini, N.E. L'Insalata, F. Rossi, and L. Fanucci, "VLSI design of a high-throughput multi-rate decoder for structured LDPC codes," in Proc. 8th EuroMicro Conf. on Digital System Design, pp. 202-209, Aug. 2005
- [4] J. Lu and J.M.F. Moura, "Structured LDPC codes for high-density recording: large girth and low error floor," IEEE Trans. Magnetics, vol. 42, no. 2, pp. 208-213, Feb. 2006.
- [5] Todd K. Moon, "Error correction coding- Mathematical Methods and Algorithms," WILEY- INDIA,2005 page- 653
- [6] T. Richardson, A. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density paritycheck codes," IEEE Trans. Inf. Theory, vol. 47, pp. 619- 637, Feb. 2001.
- [7] N. Wiberg. Codes and decoding on general graphs. PhD thesis, Linköping University, 1996. Sweden.
- [8] J. Chen, A. Dholakia, E. Eleftheriou, M.P.C. Fossorier, and X. Hu, "Reduced-complexity decoding of LDPC codes," IEEE Trans. Comms., vol. 53, no. 8, pp. 1288- 1299, Aug. 2005.
- [9] J. Zhao, F. Zarkeshvari, and A.H. Banihashemi, "On implementation of min-sum algorithm and its modifications for decoding low-density Parity-check (LDPC) codes," IEEE Trans. Comms., vol. 53, no. 4, pp. 549-554, Apr.
- [10] J. Zhang, M. Fossorier, and D. Gu, "Two-dimensional correction for min-sum decoding of irregular LDPC codes," IEEE Comm. Lett., vol. 10, pp. 180-182, Mar. 2006.

- [11] J. Chen, R.M. Tanner, C. Jones, and Y. Li, "Improved min-sum decoding algorithms for irregular LDPC codes", IEEE ISIT 2005 Proc., pp. 449-453, Sep. 2005.
- [12] M.P.C.Fossorier, M. Mihaljevic, and H.Imai, Reduced Complexity Iterative Decoding of Low-Density Parity Check Codes Based on Belief Propagation, IEEE Trans. on Comm. May 1999, vol. 47, no. 5, pp. 673-680.
- [13] Han, Wei; Huang, Jianguo; Fangfei Wu; , "A modified Min-Sum algorithm for low-density parity-check codes," Wireless Communications, Networking and Information Security (WCNIS), 2010 IEEE International Conference on , vol., no., pp.449-451, 25-27 June 2010.
- [14] Xiaofu Wu; Yue Song; Long Cui; Ming Jiang; Chunming Zhao; , "Adaptive-normalized min-sum algorithm," Future Computer and Communication (ICFCC), 2010 2nd International Conference on , vol.2, no., pp.V2-661-V2-663, 21-24 May 2010
- [15] Savin, V.; , "Self-corrected Min-Sum decoding of LDPC codes," Information Theory, 2008. ISIT 2008. IEEE International Symposium on , vol., no., pp.146-150, 6-11 July 2008 doi: 10.1109/ISIT.2008.4594965
- [16] J. Chen and M. P. Fossorier. Near optimum universal belief propagation based decoding of low density parity check codes. IEEE Trans. on Comm., 50(3):406-414, 2002.
- [17] Sarah J.Johnson, "Iterative Error Correction: Turbo, Low-Density Parity-Check and Repeat-Accumulate Codes", Cambridge University Press, 2010 , page : 64-69
- [18] Moon, T.K. (2005) "Error Correction Coding. Mathematical Methods and Algorithms", Wiley Inter-science, ISBN 0-471-64800-0
- [19] Juntan Zhang; Fossorier, M.; Daqing Gu; Jinyun Zhang; , "Improved min-sum decoding of LDPC codes using 2-dimensional normalization," Global Telecommunications Conference, 2005. GLOBECOM '05. IEEE , vol.3, no., pp. 6 pp., 28 Nov.-2 Dec. 2005.
- [20] Hrishikesh Sharma Sachin Patkar. "FPGA Design for Decoder of Projective Geometry(PG)-based Low Density Parity Check(LDPC) Codes".

AUTHORS PROFILE

Mohammad Rakibul Islam is currently working as an Associate Professor in the Department of Electrical and Electronic Engineering at Islamic University of Technology, Bangladesh. He received the B.Sc.Engg. and M.Sc.Engg. degree in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology (BUET), Bangladesh in 1998 and 2004 respectively. He also received MBA degree in Marketing from the Institute of Business Administration (IBA) under the University of Dhaka. He received his PhD degree in 2010 from Kyung Hee University, SouthKorea. His research interests include cooperative technique for wireless sensor networks, LDPC and QC-LDPC codes, secrecy capacity and other wireless applications.

Dewan Siam Shafullah received his Bachelor of Science in Engineering from Department of Electrical and Electronic Engineering from Islamic University of Technology (IUT), Bangladesh in the year of 2011. Email: siamshafullah@yahoo.com

Muhammad Mostafa Amir Faisal achieved his Bachelor of Science in Engineering from Department of Electrical and Electronic Engineering from Islamic University of Technology (IUT), Bangladesh in the year of 2011. Email: oranta68@yahoo.com

Imran Rahman obtained B.Sc.Engg. from Islamic University of Technology (IUT), Bangladesh in the year of 2011. Email: imran.iutoic@gmail.com

A New Approach of Digital Forensic Model for Digital Forensic Investigation

Inikpi O. Ademu, Dr Chris O. Imafidon, Dr David S. Preston

Dept. of Architecture,
Computing and Engineering
University of East London
London, United Kingdom

Abstract—The research introduces a structured and consistent approach for digital forensic investigation. Digital forensic science provides tools, techniques and scientifically proven methods that can be used to acquire and analyze digital evidence. The digital forensic investigation must be retrieved to obtain the evidence that will be accepted in the court. This research focuses on a structured and consistent approach to digital forensic investigation. This research aims at identifying activities that facilitate and improves digital forensic investigation process. Existing digital forensic framework will be reviewed and then the analysis will be compiled. The result from the evaluation will produce a new model to improve the whole investigation process.

Keywords – Case Relevance; Exploratory Testing; Automated Collection; Pre-Analysis; Post-Analysis; Evidence Reliability.

I. INTRODUCTION

The majority of organization relies deeply on digital devices and the internet to operate and improve their business, and these businesses depend on the digital devices to process, store and recover data. A large amount of information is produced, accumulated, and distributed via electronic means. Recent study demonstrates that in 2008, 98% of all document created in organization were created electronically (Sommer 2009). According to Healy (2008) approximately 85% of 66 million U.S. dollars was lost by organizations due to digital related crime in 2007. Panda labs (2009) show that in 2008, Ehud Tenenbaum was extradited from Canada on suspicion of stealing \$1.5million from Canadian bank through stolen credentials and infiltrated computers. Williams (2009) states on cybercrime report, a complex online fraud which scammed over £1 million pounds from taxpayers in 2009.

This research focuses on a structured and consistent approach to digital forensic investigation procedures. The research questions for the research are formulated with the aim to map out a structured and consistent approach and guideline for digital forensic investigation. This research focuses on identifying activities that facilitate digital forensic investigation, emphasizing on what digital crimes are and describing the shortcomings of current models of digital forensic investigation.

II. BACKGROUND AND RELATED WORK

Nikkel (2006) defined digital forensic as the use of scientifically derived and proven methods toward the

identification, preservation, collection, validation, analysis, interpretation, documentation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations. The term digital forensics comprises a wide range of computer activity. Not just evidence from computer, e.g. disk drive and computer memory, but including all sorts of generic media, cell phones, memory sticks, PDA's, network traffic etc. The methodologies from physical forensics are adopted into digital forensics, specific forensic software is created, and comprehensive knowledge is obtained by digital forensic specialist to defeat digital criminality.

A. Digital Evidence and its Characteristics

Carrier and Spafford (2006) defined digital evidence as a digital data that supports or refutes a hypothesis about digital events or the state of digital data. This definition includes evidence that may not be capable of being entered into a court of law, but may have investigative value, this definition is in agreement to Nikkel, (2006) definition that states, digital evidence as a data that support theory about digital events.

Evidence can be gathered from theft of or destruction of intellectual property, fraud or anything else criminally related to the use of a digital devices. Evidence which is also referred to as digital evidence is any data that can provide a significant link between the cause of the crime and the victim (Perumal, 2009).

B. Characteristics of digital evidence

- Digital evidence is by nature fragile. It can be altered, damaged or destroyed by improper handling or improper examination. It is easily copied and modified, and not easily kept in its original state, precaution should be taken to document, collect, preserve and examine digital evidence (Carrier, 2003)
- Digital evidence is a data of investigative value that is stored on or transmitted by a digital device. Therefore digital evidence is hidden evidence in the same way that Deoxyribonucleic Acid (DNA) or fingerprint evidence is hidden. In its natural state, digital evidence cannot be known by the content in the physical object

that holds such evidence. Investigative reports may be required to explain the examination process and any limitation (Pollitt, 2007).

C. Digital Devices types



Figure 1: Difference examples of Digital Devices

III. EXISTING DIGITAL FORENSIC INVESTIGATION MODELS

A. The Digital Forensic Research Workshops (DFRWS) 2001

The first DFRWS was held in Utica, New York (2001). The goal of the workshop was to provide a forum for a newly formed community of academics and practitioners to share their knowledge on digital forensic science. The audience was military, civilian, and law enforcement professionals who use forensic techniques to uncover evidence from digital sources. The group created a consensus document that drew out the state of digital forensics at that time. The group agreed and among their conclusions was that digital forensic was a process with some agreed steps. They outline processes such as identification, preservation, collection, examination, analysis, presentation and decision. (Palmer 2001). As shown in figure 4 below the grey boxes at the top of their matrix were identified by the group as fundamental processes, although many will debate the forensic nature of each step of the process. This can be called a comprehensive or an enhanced model of the DOJ model as mentioned above because it was able to cover stages that were not covered in any previous model, such as presentation stage. The main advantage of DFRWS is that it is the first large-scale organization that is lead by academia rather than law enforcement, this is a good direction because it will help define and focus the direction of the scientific community towards the challenge of digital forensic, but the DFRWS model is just a basis for future work.

B. The Forensic Process Model (2001)

According to Ashcroft (2001) the U.S National Institute of Justice (NIJ) published a process model in the Electronic Crime Scene Investigation. The document serves as a guide for the first responders. The guide is intended for use by law enforcement and other responders who have the responsibility for protecting an electronic crime scene and for the recognition,

collection and preservation of digital evidence. The forensic process consists of four phases such as:

- Collection: This involves the search for, recognition of, collection of, and documentation of electronic evidence.
- Examination: The examination process helps to make the evidence visible and explain its origin and significance. It includes revealing hidden and obscured information and the relevant documentation.
- Analysis: This involves studying the product of the examination for its importance and probative value of the case.
- Reporting: This is writing a report, outlining the examination process and information gotten from the whole investigation.

C. Abstract Digital Forensic Model (2002)

Reith, Carr and Gunsch (2002) examined a number of published models/framework for digital forensics. The basis of this model is using the ideas from traditional (physical) forensic evidence collection strategy as practiced by law enforcement (e.g. FBI). The authors argued that the proposed model can be term as an enhancement of the DFRWS model since it is inspired from it. The model involves nine components such as:

- Identification – it recognises an incident from indicators and determines its type. This component is important because it impacts other steps but it is not explicit within the field of forensic.
- Preparation – it involves the preparation of tools, techniques, search warrants and monitoring authorisation and management support.
- Approach strategy – formulating procedures and approach to use in order to maximize the collection of untainted evidence while minimizing the impact to the victim
- Preservation – it involves the isolation, securing and preserving the state of physical and digital evidence
- Collection – This is to record the physical scene and duplicate digital evidence using standardized and accepted procedures
- Examination – An in-depth systematic search of evidence relating to the suspected crime. This focuses on identifying and locating potential evidence.
- Analysis – This determines importance and probative value to the case of the examined product
- Presentation - Summary and explanation of conclusion
- Returning Evidence – Physical and digital property returned to proper owner

D. The Integrated Digital Investigation Process Model (IDIP) 2003

Carrier and Spafford (2003) proposed a model, which the authors provide a review of previous work and then map the digital investigative process to the physical investigation process. The model known as the Integrated Digital Investigation Process was organized into five groups consisting of 17 phases.

E. Enhanced Digital Investigation Process (2004)

Baryamueeba and Tushaba (2004) suggested a modification to Carrier and Spafford's Integrated Digital Investigation Model (2003). In the model, the authors described two additional phases which are trace back and dynamite which seek to separate the investigation into primary crime scene (computer) and secondary crime scene (the physical crime scene). The goal is to reconstruct two crime scenes to avoid inconsistencies.

F. Extended model of cyber crime investigation

Ciardhuain (2004) argues that the existing models are general models of cybercrime investigation that concentrate only on processing of evidence in cybercrime investigation. The model shown provides a good basis for understanding the process of investigation and captures most of the information flows. Even though the model was generic, it concentrated on the management aspect.

G. Case-Relevance Information Investigation (2005)

Ruibin, Yun and Gaertner (2005) identified the need of computer intelligence technology to the current computer forensic framework. The authors explained that computer intelligence is expected to offer more assistance in the investigation procedures and better knowledge reuse within and across multiple cases and sharing. First concept that was introduced by the authors is the notion of Seek Knowledge which is the investigative clues which drive the analysis of data. Another concept described by the authors is the notion of Case-Relevance. They used this notion to describe the distinctions between computer security and forensics even defining degrees of case relevance.

H. Digital Forensic Model based on Malaysian Investigation Process (2009)

Perumal (2009) proposed a model that clearly defines that the investigation process will lead into a better prosecution as the very most important stages such as live data acquisition and static data acquisition has been included in the model to focus on fragile evidence.

I. The Systematic digital forensic investigation model SRDFIM (2011)

Agawal et al (2011) developed a model with the aim of helping forensic practitioners and organizations for setting up appropriate policies and procedures in a systematic manner. The proposed model in this paper explores the different processes involved in the investigation of cybercrime and cyber fraud in the form of an eleven stage model. The model focuses on investigation cases of computer frauds and cyber-crimes. The application of the model is limited to computer frauds and cyber-crimes.

IV. PROPOSED MODEL

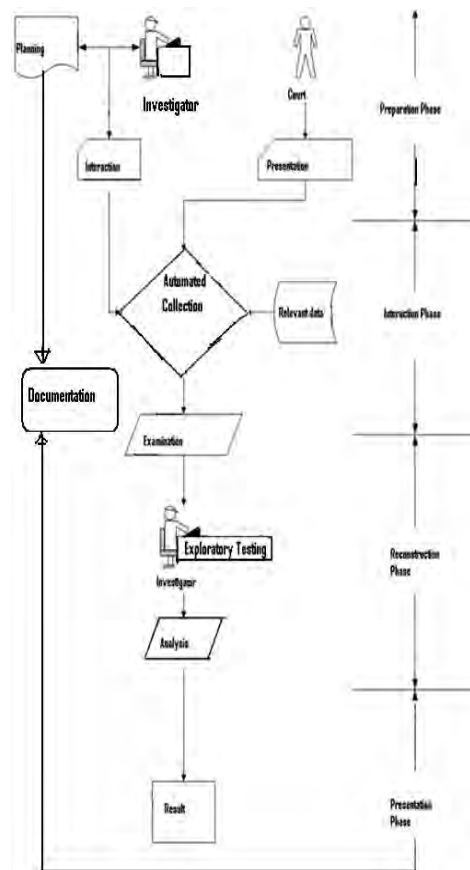


Figure 2: Proposed digital forensic investigation Model

In the proposed model the digital forensic investigation process will be generalised into 4 tier iterative approach. The entire digital forensic investigation process can be conceptualized as occurring iteratively in four different phases. The first tier which is the preparation or inception phase occur over the course of an investigation from assessment to final presentation phase. The first tier will have 4 rules for digital forensic investigation which involves preparation, identification, authorisation and communication. The second tier will have rules such as collection, preservation and documentation, the third tier will have rules consisting examination, exploratory testing, and analysis, the 4th tier which is the presentation phase have rules such as result, review and report.

J. Advantages and Disadvantages of Proposed Model

The model has the advantages obtained from existing model and then expands its scope and provides more advantages. A structured and consistent framework is vital to the development of digital forensic investigation and the identification of areas in which research and development are needed.

The model identifies the need for interaction. Investigator should have consistent interaction with all resources for carrying out the investigation.

Knowing the need of the client/victim and determining to meet the need is important. Better case goal can be defined. Optimal interaction with tools used by investigator is very important. Tools need to be used by people who know how to use them properly following a methodology that meets the legal requirement associated with the particular jurisdiction.

Another advantage of the model is exploratory testing. Investigators need to have the patience, to stay on the target and have to learn any new techniques while performing an investigation. Very little testing has been formalized in this field for the specific need of digital forensic, investigators wishing to be prudent should undertake their own testing methods and this should be a normal part of the process used in preparing for legal matters and this should also meet the legal requirement of the jurisdiction

The model can also help capture the expertise of investigation as a basis to the development of advanced tools incorporating techniques such as automated digital evidence collection.

Generality of the model is not explicit. It must be applied in the context of a crime before it will be possible to make clear the details of the process.

IV. CONCLUSION

Digital evidence must be admissible, precise, authenticated and accurate in order to be accepted in the court. Digital evidence is fragile in nature and they must be handled properly and carefully. A detailed digital forensic procedure provides important assistance to forensic investigators in gathering evidence admissible in the court of law.

In completing the proposed research, I will learn how apply the proposed system to digital forensic investigation. Bearing this in mind, my expected result, are firstly, to develop a model from relevant domains and bodies of theory of digital forensic and secondly a set of implementable guidelines of digital forensic investigation will be identified.

The digital forensic community needs a structured framework for rapid development of standard operational procedures that can be peer – reviewed and tested effectively and validated quickly.

Digital forensic practitioners can benefit from the iterative structure proposed in this research to build forensically sound case and also for the development of consistent and simplified forensic guides on digital forensic investigation that can be a guideline for standard operational procedure and a model for developing future technology in digital forensic investigation.

REFERENCES

- [1] Agrawal, A. Gupta, M. Gupta, S. Gupta, C. (2011) Systematic digital forensic investigation model Vol. 5 (1) Available (online): <http://www.cscjournals.org/csc/manuscript/Journals/IJCSS/volume5/Issue1/IJCSS-438.pdf> Accessed on 30th June 2011
- [2] Ashcroft, J (2001) Electronic Crime Scene Investigation: A guide for first responders Available (online): <https://www.ncjrs.gov/pdffiles1/nij/187736.pdf> Access on 20th October 2011
- [3] Baryamureeba, V. Tushabe, F. (2004) The Enhanced digital investigation process (2004) Available (online): <http://www.dfrws.org/2004/bios/day1/tushabeEIDIP.pdf> Accessed on 15th June 2011
- [4] Carrier, B. Spafford, H. (2006), Getting physical with digital forensic process Vol. 2 (2) Available (online): <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.76..pdf> Accessed on 20th August 2011
- [5] Carrier, B. (2003) Defining digital forensic examination and analysis tools using abstraction layers Vol. 1 (4) Available (online): <http://www.cerias.purdue.edu/homes/carrier/forensics> Accessed on 20th September 2011
- [6] Ciardhuain, S. (2004) An extended model of cybercrime investigation Accessed on 20th October 2011 Available (online): www.ijde.org/citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.80...A cessed on 11th August 2011
- [7] Healy, L. (2008) Increasing the Likelihood of admissible electronic evidence: Digital log Handling excellence and a forensically aware corporate culture Available (online): <http://www.emich.edu/ia/pdf/phdresearch/Increasing%20the%20Likelihood%20of%20Admissible%20Electronic%20Evidence,%20Larry%20Healy%20COT%20704.pdf>. Accessed on 20th August 2011
- [8] Nikkel, B. (2006) the role of digital forensic with a corporate organisation Available (online): www.digitalforensics.ch/nikkel/06a.pdf Accessed on 25th February 2010
- [9] Palmer, G. (2001) a road map to digital forensic research Available (online): <http://www.dfrws.org/2001/dfrws-rm-final.pdf> Accessed on 25th October 2011 Panda labs Annual Report (2009) Available (online): http://www.pandasecurity.com/img/enc/Annual_Report_Pandalabs2009.pdf Accessed 16th August 2011
- [10] Perumal, S. (2009) Digital forensic model based on Malaysian investigation process Vol. 9 (8) Available (online): http://paper.ijcsns.org/07_book/200908/20080805.pdf Accessed on 7th August 2011
- [11] Pollitt, M. (2007) An Ad Hoc Review of Digital Forensic Models, Vol. 10(12) Available (Online): <http://www.ieeexplore.ieee.org/ie15/4155337/4155338/04155349.pdf?> Accessed on the 17th September 2011
- [12] Reith, M. Carr. C. Gunsch, G. (2002) an examination of digital forensic model. Department of Electrical and Computer Engineering Air force institute of technology. Wright-Patterson. Available (Online): <http://www.utica.edu/academic/institutes/ecii/ijde/articles.cfm?action> Accessed on the 7th October 2011.
- [13] Ruibin, G. Garrtner, M. (2005) Case-Relevance Information Investigation: Binding Computer Intelligence to the Current Computer Forensic Framework. Vol. 4(1) Available (Online): <http://www.utica.edu/academic/institutes/ecii/publications/articles/B4A6A102-A93D-85B1-95C575D5E35F3764.pdf> Accessed 15th September 2011

A Data Mining Approach for the Prediction of Hepatitis C Virus protease Cleavage Sites

Ahmed mohamed samir ali gamal eldin

Bio-informatics
Helwan University
Cairo, Egypt

Abstract— Summary: Several papers have been published about the prediction of hepatitis C virus (HCV) polyprotein cleavage sites, using symbolic and non-symbolic machine learning techniques. The published papers achieved different Levels of prediction accuracy. the achieved results depends on the used technique and the availability of adequate and accurate HCV polyprotein sequences with known cleavage sites. We tried here to achieve more accurate prediction results, and more Informative knowledge about the HCV protein cleavage sites using Decision tree algorithm. There are several factors that can affect the overall prediction accuracy. One of the most important factors is the availability of acceptable and accurate HCV polyproteins sequences with known cleavage sites. We collected latest accurate data sets to build the prediction model. Also we collected another dataset for the model testing.

Motivation: Hepatitis C virus is a global health problem affecting a significant portion of the world's population. The World Health Organization estimated that in 1999; 170 million hepatitis C virus (HCV) carriers were present worldwide, with 3 to 4 million new cases per year. Several approaches have been performed to analyze HCV life cycle to find out the important factors of the viral replication process. HCV polyprotein processing by the viral protease has a vital role in the virus replication. The prediction of HCV protease cleavage sites can help the biologists in the design of suitable viral inhibitors.

Results: The ease to use and to understand of the decision tree enabled us to create simple prediction model. We used here the latest accurate viral datasets. Decision tree achieved here acceptable prediction accuracy results. Also it generated informative knowledge about the cleavage process itself. These results can help the researchers in the development of effective viral inhibitors. Using decision tree to predict HCV protein cleavage sites achieved high prediction accuracy.

Keywords-component; HCV polyprotein; decision tree; protease; decamers

I. INTRODUCTION

Hepatitis C virus (HCV) is a virus that infects liver cells and causes liver inflammation. It is a global disease with a worldwide expanding incidence and prevalence base. Hepatitis C virus presents supremely challenging problems in view of its adaptability and its pathogenic capacity. The strategies that HCV utilizes to parasitize its hosts make it formidable enemy. Therapeutic interventions need considerable sophistication to counter its progress. It is estimated that 3–4 million people are infected with HCV each year. Some 130–170 million people

are chronically infected with HCV and at risk of developing liver cirrhosis

and/or liver cancer. More than 350 000 people die from HCV related liver diseases each year.

HCV infection is found worldwide. Countries with high rates of chronic infection are Egypt (22%), Pakistan (4.8%) and China (3.2%). these countries are attributed to unsafe injections using contaminated equipment. [1].

HCV protease cleavage sites are considered one of the most important inhibitor targets, cause of the cleavage of polyprotein Sequences plays an important role in the viral replication [2].

The prediction of the viral proteases cleavage sites will help in the development of suitable protease inhibitor. Several data mining techniques have been used in solving and analyzing several biological problems. One of the interesting problems is the analyzing of HCV life cycle, using Data mining techniques to find useful knowledge which may help the biologist to develop suitable HCV vaccine. Many data mining techniques have been used to analyze different viral proteases cleave sites. For example artificial neural network has been used to predict both Human immunodeficiency virus (HIV) and HCV proteases cleavage sites and achieved high prediction accuracy [3-5]. Finding more accurate and simpler prediction model is considered a challenging point.

Decision tree is one of the most common data mining techniques. It has been used in analyzing and solving several classification problems. Decision tree has a great advantage which its ability to provide us with informative rules about the classification problem itself. The biologists and the researchers can use these rules to understand the cleavage

Process characteristic. In spite of that decision tree does not have prediction accuracies than the other classification techniques, but its ease to understand and also its informative rules make it an interesting method. Decision tree prediction results depends on the availability of the datasets which it will train the classification model. Decision tree has been used in the prediction of HCV protease cleavage sites, but it did not achieve an acceptable prediction accuracies cause of the lake of accurate cleaved sequences [6]. We tried

Here to collect and find more accurate HCV cleaved sequences to build a decision tree to predict the proteases cleavage sites.

II. SYSTEM AND METHODS

A. Viral protease cleavage process

The cleavage process of the protein is look like the 'Lock and key' model where a sequence of amino acids fits as A key to the active site in the protease, which in the HCV protease Case is estimated to be ten residues long. The protease active site pockets are denoted by S (Schechter and Berger, 1967) [7].

$S = S5, S4, S3, S2, S1, S1', S2', S3', S4', S5'$

Corresponding to residues P in the peptide

$P = P5, P4, P3, P2, P1, P1', P2', P3', P4', P5'$

The scissile bond is located between positions P1 and P1', and Pi can take on any one of the following 20 amino acid values {A, R, N, D, C, Q, E, G, H, I, L, K, M, F, P, S, T, W, Y, V}. There are 2010 possible values for string P. If the amino acids in P (the 'key') fit the positions in S (the 'lock'), then the protease will cleave the decamer (ten amino acids) between positions P1 and P1'. The goal of the decision tree model to learn the 'lock and key' rules, from the available datasets.

B. Date representation

HCV protein sequences are represented as a long chain of letters. Each letter represents one amino acid. We interested in the 10 amino acids where the HCV protease can cleave it. There is a poplar technique used by the previous researches to generate non-cleaved sequences [6]. It depends on considering the regions between known cleaved sequences as a non-cleaved.

C. Building the classification model

We used here one of the most common used classification algorithms which is the decision tree algorithm. We will summarize basic concepts of the decision trees and its advantages over the other classification methods. A decision tree is a tree in which each branch node represents a choice between a number of alternatives, and each leaf node represents a decision. Decision tree are commonly used for gaining information for the purpose of Decision -making. Decision tree starts with a root node on which it is for users to take actions. From this node, users split each node recursively according to Decision tree learning algorithm. The final result is a decision tree in which each branch represents a possible scenario of decision and its outcome [8]. The following is a summary of the important characteristics of decision

- Decision tree induction is a nonparametric approach for building classification models. In other words, it does not require any prior assumptions regarding the type of probability distributions satisfied by the class and other attributes.
- Techniques developed for constructing decision trees are computationally inexpensive, making it possible to quickly construct models even when the training set size is very large. Furthermore, once a decision tree has been built, classifying a test record is extremely fast.
- Decision trees, especially smaller-sized trees, are relatively easy to interpret. The accuracies of the trees

are also comparable to other Classification techniques for many simple data sets.

- Decision trees provide an expressive representation for learning discrete valued functions.
- Decision tree algorithms are quite robust to the presence of noise, especially when methods for avoiding overfitting.

For the HCV protein cleave sites prediction problem. We used the decision tree model with Gini index splitting rule [9]. Each sample in the training dataset was consisting of 11 items 10 items represent the amino acids where the protease can cleave it. The last item represents the class label of the amino acids sample. In our problem we have two classes cleavage 'positive' or non-cleavage 'negative'.

D. Data collection

The process of collecting enough and accurate HCV cleaved decamers, is the core of our research. We searched a lot of the published papers that have discussed HCV polyprotein analysis. Also we contacted a lot of researchers interested in this area. The availability of the online protein databases provided us with some accurate and valid HCV polyprotein sequences for the training and testing our model. To generate more non- cleaved 'negative' sequences we used the technique which has been used by the previous researchers as we mentioned in the previous section.

There are several conflicts and uncertainties in the data which have been used in the previous published papers. We tried to found the most recent and accurate samples to build the prediction model. We used the last accurate datasets used [10-18] in previous work. The collected datasets are divided into two parts:

- Training dataset
- out of sample or testing dataset

We collected 939 decamers as training dataset 199 as cleaved 'positive' samples and 706 as non-cleaved 'negative' samples. We collected three out of samples dataset to the proposed model [19]:

- Four proteins from the TLR3 pathway were used for another test data set: IκB kinase ε (IKKε) [GenBank: AAC51216]; TRAF family member-associated NF-κB activator-binding kinase 1 (TBK1) [GenBank: NP_037386]; Toll-like receptor 3 (TLR3) [GenBank: NP_003256]; and Toll-IL-1 receptor domain containing adaptor inducing IFN-β (TRIF or TICAM-1) [GenBank: BAC55579].the four proteins created dataset contains 2806 samples of which two are reported as cleaved samples by HCV protease enzyme[20].
- There are 69 samples reported in vivo as cleaved samples[19].

We used the same datasets for training and testing which have been used by Thorsteinn Rögnvaldsson et al [19]. They collected a new datasets rather than the previous datasets used

by the other researchers, which contains a lot of conflict and uncertainties as we mentioned before.

III. RESULTS AND DISCUSSION

We implemented the decision tree using classification and regression tree (CART) Mat lab toolbox with GINI index as spitting criteria. The training dataset was consisting of 939 samples.199 as cleaved sample and 740 as non-cleaved samples. Each sample was consisting of 10 amino acids where the HCV protease can cleave.

We used ten-fold cross validation to be able to evaluate the overall performance of the prediction model. For the training dataset we got Prediction accuracy 99 %. Also we got 98% as Sensitivity and Specificity as 99%. Table I show the confusion matrix for the training data.

After apply the ten-fold cross validation got overall accuracy 96% and we got Sensitivity is 95.5% and the model Specificity 98.6%. Table II shows the average achieved confusion matrix for the tenfold cross validation.

We applied our model on the out of samples dataset. For the first test set which is consist of 2806 (2 cleaved and the remaining are non-cleaved samples) sample. Our model successfully predicted one of the cleaved samples. But it got 89 as false positive or false cleaved samples.

For the 69 in vivo cleaved samples our model successfully predicted 59 of the 69 as cleaved samples.

Using the decision tree as a classification model has achieved an overall prediction accuracy 96% which can be considered as an acceptable results, if we compared the presented model with the other techniques that achieved the

Highest prediction accuracy like support vector machine (SVM) [5]. We can find that our results are comparable with SVM which achieved 97% as overall prediction accuracy.

The presented work is a try to achieve more accurate prediction accuracy using easy and simple classification technique like the decision.

IV. CONCLUSIONS AND FUTURE WORK

The prediction of HCV polyproetin cleavage sites, using Decision tree, has achieved acceptable prediction accuracies. The achieved results are not the best, but the created rules by the decision tree prediction model make the achieved results more informative. In the future work we can add more factors like the amino acids secondary structure as training attribute to find out its effect in the overall prediction accuracy. Also we can enhance the decision tree prediction results by using the ensembles of decision tree technique which can enhance the prediction results of the proposed model.

TABLE I. THE CONFUSION MATRIX FOR THE TRAINING DATA

	Non cleavage	cleavage	Total
None cleavage	735	5	740
Cleavage	4	195	199

Total	739	200	939
-------	-----	-----	-----

TABLE II. THE CONFUSION MATRIX FOR 10-FOLD CROSS VALIDATION

	Non cleavage	cleavage	Total
None cleavage	730	10	740
Cleavage	9	190	199
Total	739	200	939

REFERENCES

- [1] World healt oragnization Media centre. "Hepatitis C ." <http://www.who.int>. 2011. 5 October 2011 <http://www.who.int/mediacentre/factsheets/fs164/en/>
- [2] Sarah Welbourn and Arnim Pause," The Hepatitis C Virus NS2/3 Protease," Molecular Biology (2007), in press
- [3] T. Rognvaldsson, Liwen You , "No Algorithm Beats the Simple Perceptron on HIV Protease Function Prediction," unpublished .
- [4] Thompson, T., Chou, K, and Zheng, C. , "Neural network prediction of the HIV-1protease cleavage sites". Journal of Theoretical Biology (1995)177, 369-379," inpress .
- [5] T Cai, Y.-D. and Chou, K.-C., "Artificial neural network model for predicting HIV protease cleavage sites in protein," Advances in Engineering Software (1998) 29, 119-128 .
- [6] Ajit Narayanan, Xikun Wu and Z. Rong Yang," Mining viral protease data to extract cleavage knowledge," Bioinformatics (2002) 18 (suppl1): S5-S13,In press
- [7] T. Rognvaldsson, Liwen You , "Why neural networks should not be used for HIV-1 protease cleavage site prediction," Bioinformatics (2004), in press .
- [8] W. Peng, J. Chen and Haiping Zhou," An Implementation of ID3 Decision Tree Learning Algorithm," unpublished
- [9] Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984),"Classification and regression trees." Wadsworth, Belmont. Refrence
- [10] Jarman IH, Etschells TA, Martin JD, Lisboa PJ (2008) an integrated framework for risk profiling of breast cancer patients following surgery. Artificial Intelligence in Medicine, 42:165-188
- [11] Grakoui A, McCourt DW, Wychowski C, Feinstone SM, Rice CM: Characterization of the hepatitis C virus-encoded serine proteinase: determination of proteinase-dependent polyprotein cleavage sites. Journal of Virology 1993, 67:2832-2843.
- [12] Leinbach SS, Bhat RA, Xia SM, Hum WT, Stauffer B, Davis AR, Hung PP, Mizutani S: Substrate specificity of the NS3 serine proteinase of hepatitis C virus as determined by mutagenesis at the S3/NS4A junction. Virology 1994, 204:163-169.
- [13] Kolykhalov AA, Agapov EV, Rice CM: Specificity of the hepatitis C virus NS3 serine protease: effects of substitutions at the 3/ 4A, 4A/4B, 4B/5A, and 5A/5B cleavage sites on polyprotein processing. Journal of Virology 1994, 68:7525-7533.
- [14] Bartenschlager R, Ahlborn-Laake L, Yasargil K, Mous J, Jacobsen H: Substrate determinants for cleavage in cis and in trans by the hepatitis C virus NS3 proteinase. Journal of Virology 1995, 69:198-205.
- [15] Urbani A, Bianchi E, Narjes F, Tramontano A, Francesco RD, Steinkühler C, Pessi A: Substrate specificity of the hepatitis C virus serine protease (NS3). The Journal of Biological Chemistry 1997, 272:9204-9209.
- [16] Zhang R, Durkin J, Windsor WT, McNemar C, Ramanathan L, Le HV: Probing the substrate specificity of hepatitis C virus NS3 serine protease by using synthetic peptides. Journal of Virology 1997, 71:6208-6213.
- [17] Kwong AD, Kim JL, Rao G, Lipovsek D, Raybuck SA: Hepatitis C virus NS3/4A protease. Antiviral Research 1998, 40:1-18.
- [18] Attwood MR, Bennett JM, Campbell AD, Canning GGM, Carr MG, Conway E, Dunsdon RM, Greening JR, Jones PS, Kay PB, Handa BK,

- [19] Hurst DN, Jennings NS, Jordan S, Keech E, O'Brien MA, Overton HA, Wilkinson TCI, Wilson FX: The design and synthesis of potent inhibitors of hepatitis C virus NS3-4A proteinase. *Antiviral Chemistry & Chemotherapy* 1999, 10:259-273.
- [20] T. Rögnvaldsson, T. A Etchells, L. You, "How to find simple and accurate rules for viral protease cleavage specificities," *BMC Bioinformatics* 2009, in press
- [21] Li K, Foy E, Ferreon JC, Nakamura M, Ferreon ACM, Ikeda M, Ray SC, "Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF", .
- [22] Proceedings of the National Academy of Sciences of the United States Of America 2005, in press.

Enhancing Business Intelligence in a Smarter Computing Environment through Cost Analysis

Saurabh Kacker, Vandana Choudhary, Tanupriya Choudhury, Vasudha Vashisht
Department of Computer Science
Lingaya's University
Faridabad, India

Abstract—The paper aims at improving Business Intelligence in a Smarter Computing Environment through Cost Analysis. Smarter Computing is a new approach to designing IT infrastructures to create new opportunities like creating new business models, find new ways of delivering technology-based services, and generate new insights from IT to fuel innovation and dramatically improve the economics of IT. The paper looks at various performance metrics to lower the cost of implementing Business intelligence in a smarter computing environment, to generate a cost efficient system. To ensure it, smarter services are deployed with business strategy. The working principle is based on workloads optimizations and their corresponding performance metrics like value metrics, Advanced Data Capabilities and Virtualizations so as to decrease the total IT cost.

Keywords— Smarter Computing; Business Intelligence; Cost Analysis; Virtualizations; Advanced Data Capabilities; Value Metrics.

I. INTRODUCTION

We can see dramatic shifts as our planet becomes smarter. These shifts are changing the way the world works. Cities are becoming smarter by transforming traffic systems, water systems, security—every possible form of municipal infrastructure. Business process is evolving across every industry—banking, trading, manufacturing. And we're seeing changes in the way people live, enjoying advancements ranging from reduced congestion and pollution to new ways to communicate and collaborate. Every aspect of life is benefited from the instrumentation, interconnection and infusion of intelligence into the systems of the world [11].

Nothing is changing more than information technology: the way it's accessed, the way it's applied, and the way it's architected [11]. Change is the name of the game – particularly in Information Technology, it starts with the ever-changing expectations of everyone in the chain: the providers of IT, the partakers and the consumers. To a large extent this is driven by evolving business landscape, stiffer competition and the threat of becoming too outdated. Customer demands for top value products and services are escalating – with which IT infrastructure and systems simply need to keep up. The opportunities for innovation have never been greater.

By thinking differently about computing, the leaders have addressed the IT conundrum—meeting exploding demand for service on a flat budget. This conundrum traps IT organizations

in a vicious cycle in which a rigid infrastructure and lack of trusted data lead to reactive decision-making. Meanwhile, attempts to overcome these challenges through additional IT investments result in a more sprawling and costly infrastructure [8].

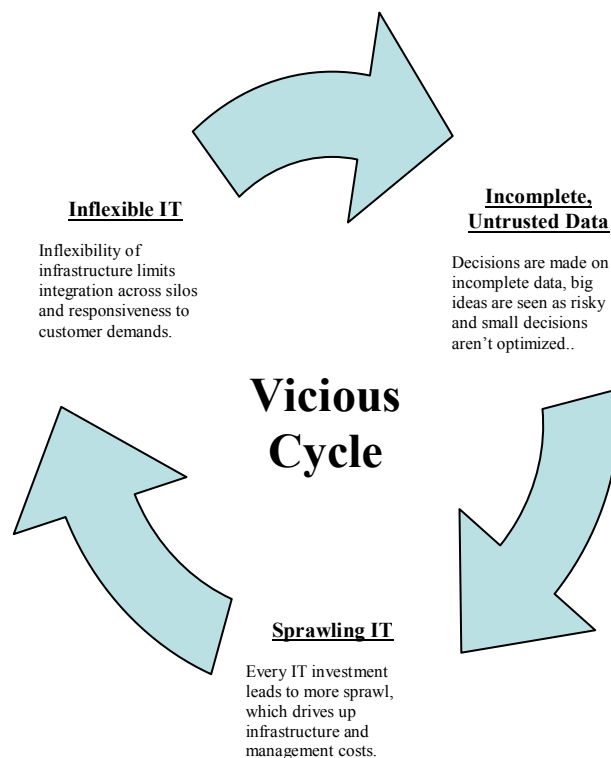


Figure 1. Current IT Business Situation

Leaders in this new era understand that the traditional approach of just adding more servers, storage, networking and other system components to meet increasing demand is no longer a sustainable model. This traditional approach ends in hardware sprawl, bloated software and labour costs, higher environmental costs, all contributing to higher total IT costs. This makes us to enter in a new era of computing known as Smarter Computing—the era of insight for discovery. Smarter Computing is taking a new approach to designing IT infrastructures to create new opportunities like creating new

business models, find new ways of delivering technology-based services, and generate new insights from IT to fuel innovation and dramatically improve the economics of IT. A smarter approach to computing makes it possible to meet increasing demand and to support innovation while managing to stay within nearly flat IT budgets. Optimized Systems for superior economics and Cloud to reinvent business processes and drive innovation. Any enterprise can enter this new era by architecting an IT infrastructure that is designed for data, tuned to the task and managed in the cloud. Thus Smarter Computing will improve the conundrum that is trapping us into a vicious cycle and making a virtuous environment for IT leaders [8].

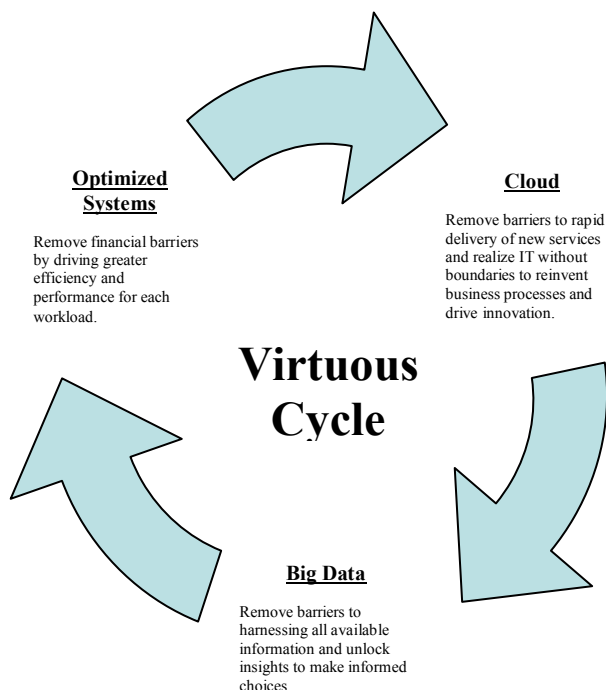


Figure 2. Reversing the Current IT Business situation through Smarter Computing

II. PROBLEM FORMULATION

Recently the research community has believed that Business Intelligence has become choice of midsize and large size companies that want to improve business processes, discover new opportunities, and edge out the competition. Business Intelligence gives you the information when you need it, in the format you need. By integrating data from across your enterprise and delivering self-service reporting and analysis, IT spends less time responding to requests and business users spend less time looking for information. Business Intelligence tries to improve the productivity of business analysts and preserve information consistency throughout an organization. Smarter Computing is a promising environment which provides computation and storage resources as services along with efficient method to reduce efforts. Smarter computing infrastructure and related mechanisms allows the stability, equilibrium, efficient resource use and sustainability of a Business Intelligence System [4].

IT serves the strategy of the business, but keeping business objectives and IT operations in alignment is not always easy. We may have a clear business strategy mapped to detailed business processes that are ready to implement but still the execution stumbles. Why? One reason is that the information systems needed to execute the strategy are insufficient or poorly matched to the requirements. Smarter computing and service oriented architectures can mitigate the risk of such misalignments, assuming they are used in ways supportive of business strategy. Aligning business strategy and IT services is a several step process, at least at the most coarse level:

- Identifying key business objectives
- Identifying IT services needed to support those objectives
- Assessing the current state of IT services and identifying gaps between the existing set and the needed set of IT services.
- Developing a plan for reducing the gap between the existing and needed set of information services

Key business objectives may include controlling and reducing costs, enabling more rapid response to changing market conditions, improving governance of the organization, or improving the resiliency of IT operations to adverse events, such as hardware failures, loss of power, or natural disaster. Many of the services needed to support business objectives can be readily identified once the business objectives are known. Cost controls and cost reduction come with more efficient server utilization, more self- service in systems management, and reduced overhead associated with infrastructure services such as backups, high availability, and disaster recovery. The gap analysis process should take into account both technical and organizational considerations. For example, will existing hardware readily deploy in cloud architecture or will new hardware be required? Are service management practices mature enough to implement in self- service delivery systems? Is a billing or chargeback mechanism in place if a private cloud is under consideration? The first steps in creating a plan to move from the existing to the needed systems are to prioritize the gaps and identify dependencies in the process. The next step is creation of Dynamic Infrastructure. A Dynamic Infrastructure is required for delivery of smarter products and services. A dynamic infrastructure is one that is designed to enable today's instrumented and interconnected world, helping clients integrate new smarter components of their business with a flexible, secure and seamlessly managed IT infrastructure. Organizations supported by a dynamic infrastructure can deliver superior business and IT services with agility and speed, while addressing the day-to-day operational needs to improve service, reduce cost and manage risk. A first step towards enabling a dynamic infrastructure is to break down the traditional barriers that separate software development and IT operations. In most organizations today, a separation exists between software development and IT operations. The two groups use different tools, speak different languages, maintain proprietary data stores and expend few resources on integration with external entities.

Generally, this separation is born out of the requirement to enable the two functions to specialize and focus on their individual goals. For development the goal is application creation, for IT operations, the goal is service assurance. Given this distinction, specialization within development and operations makes sense. As with many undertakings, specialization underpins organizational effectiveness, creating value by allowing teams to focus on a subset of well-defined objectives.

Specialization, however, only works up to the point where work flows are contained within organizational boundaries. When focus is put on the on the larger goal of end-to-end service delivery where workflows must be coupled across teams, this specialization of tools, skills, and processes can expose the business to unwarranted risk [1].

III. ARCHITECTURAL STRATEGY

Current changes in IT and User expectations make architecting for a smarter strategic computing plan a necessity for businesses. Alignment with past technology initiative like Service Oriented Architecture (SOA) and delivery processes like Agile is important. After all, many organizations are deep into their SOA and data centre consolidation & transformation programs so any low hanging fruit that they can obtain towards a smarter computing path needs to be exposed.

However, a strong architectural governance and management process is by no means the only criteria for a smarter computing plan. Not all organizations or businesses will face the same integration, architectural or evolutionary challenges. That is why there is a resurgence of Service Level Agreements (SLA) which provides business units with performance metrics that they can monitor as part of their justification for using internal rather than external services. Business managers may look at cloud services and find the lower costs, greater control, and potential for scaling business processes compelling reasons to use cloud services. These reasons are often not enough, though. It is not sufficient for a cloud to work well today. This is why we need SLAs. SLAs are standard in IT, and it is no surprise that they are used with cloud services. Rather than focus just on the availability of a specific application, cloud SLAs may be more general and apply to capacity commitments, network infrastructure, storage infrastructure, and availability and recovery management. These SLAs are closely coupled to the infrastructure of the cloud, but the primary concern is on the business commitments cloud providers make to their customers.

Over the years, Organizations have implemented numerous architectural strategies and business driven IT systems to create an integrated business oriented computing strategy. Business processes have been reengineered over and over again to meet each new technology model of architecture change. For different industries and organizations sizes there may be a broad range of concerns and issues to consider ranging from regulatory or statutory requirements, to technology maturity and management commitment and culture.

Ability to Integrate, Automate and apply secure design and delivery practices are amongst the necessary criteria for delivering a framework that is able to:

1. Respond and deliver value quickly and pragmatically.
2. Make use of existing investments and incorporates a broad range of appropriate facilities and resources.
3. Address and protect against key security concerns and requirements.

Making decisions on the above factors help in designing a smarter computing plan for Business valued IT. The IT estate for many businesses and organizational data centres has expanded to include a wide range of infrastructure platforms from mainframes to distributed servers. Advances in software infrastructure, application technology and the arrival of service based delivery models have resulted in businesses facing new architectural models and challenges:

1. *Infrastructure*: Virtualized, hosted/service based, distributed, complex, hybrid on demands.
2. *Application*: Virtualized, distributed, packaged complex, custom, on demands, mobile.
3. *Data*: Federated, private, public, big, structured, unstructured, transient.
4. *IT Process*: Mixed (agile, iterative, formal), automated.
5. *Mobile*: Smartphone, Tablets, hotel business centres, internet cafes [7].

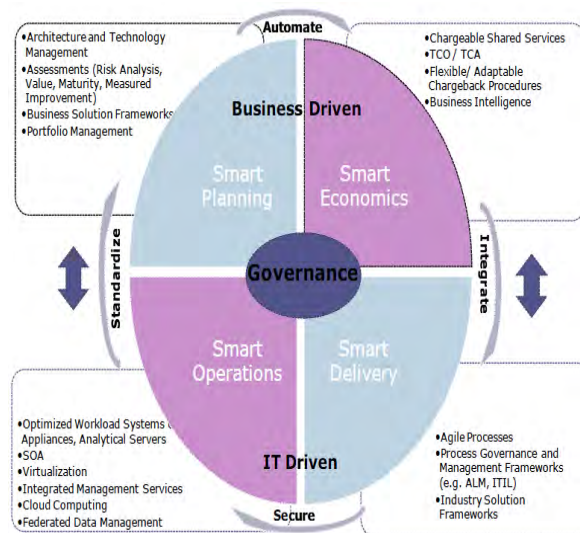


Figure 3. Integrated Strategy for Business valued IT

A Workload Optimizations

A large server farm is indistinguishable from a set of cloud servers when looking at the hardware. Servers, switches, routers, power supplies, and other components are the same.

The difference lies in how these resources are used. The servers in a typical corporate data centre prior to the advent of cloud computing were assigned to a particular department or application use. The configuration was relatively fixed and changed only when the server was upgraded, reassigned, or decommissioned. These servers were configured to do one type of operation. This makes for a reliable compute resource, but not an efficient one.

Servers with fixed configurations are less likely to have high- utilization rates. Unless there is a steady stream of jobs that fits the machine's configuration, there will be idle periods. Without proper infrastructure for rapidly deploying virtual machines, the cost of reconfiguring a server is so high that it is done only for significant long- term changes. In the cloud, the cost of switching virtual machines is low enough that idle servers can be reconfigured with different virtual machine images allowing other applications to run on the same physical server that had just been running other types of jobs.

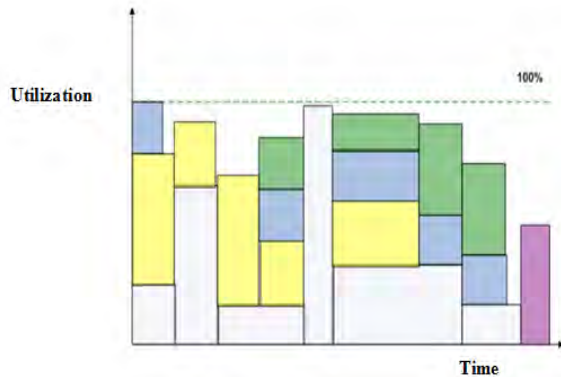


Figure 4. Cloud Server Utilization

The figure shows that in the cloud, server utilization can be significantly higher when workloads are distributed and optimized over available servers [4].

B. Virtualizations

Currently, Virtualization revolves around use of commodity server to better utilize resources. For general purpose computing such as email server, file or print server and many application server hardware utilization hardly goes above 10%. Using Virtualizations to get better utilization has reduced the physical estate substantially.

Virtualization is not just about servers. Network and storage virtualization are also used in many organizations as they look to improve throughput and get the most use out of the assets. Storage Virtualization is critical to the ability to support the huge volume of data available today. There are others benefits to the wider use of Virtualization such as backup, replication and the ability to clone mission critical servers on demand [7].

Setting up a set of virtual machines on a single server is straightforward: install a hypervisor and create virtual machine instances based on OS(s) of choice. Scaling virtualization to a large number of servers requires management software that can manage multiple hypervisor clients from a single console. Storage services also need to be virtualized so that they appear to cloud consumers to be a single storage device. Virtual machine instances in the cloud, for example, should be able to address storage space on the cloud SAN(s) without having to manage implementation Details. Ideally, the same management console that is used to control servers in the cloud will support management and administration of storage resources. Computing and storage clouds hide many of the implementation details that go into Building and maintaining a large IT infrastructure. By standardizing services, streamlining service

management, and virtualizing physical resources, cloud providers enable the technical resources needed by users to leverage cloud services. Those same users, however, also require attention to business considerations [4].

IV. COST ANALYSIS

The current system in which Business Intelligence is implemented is not flexible enough to provide result at a low cost as well as there is no load management between multiple work flows. Smarter Computing aims at transforming the economics of IT while freeing teams to focus on new innovation. To ensure smarter services are deployed with business strategy, we should focus on workloads optimization and their corresponding performance metrics so as to decrease the total IT cost [3].

A. Workload Analysis

Right now in business there are hundreds, thousands, or even more applications executing business processes. Some of these are transaction- processing systems that provide high- volume, rapid processing of orders, inquiries, reservations, or a broad array of other narrowly focused business activities. Other applications are performing batch operations, such as generating invoices, reviewing inventory levels, or performing data quality control checks on databases. Still others are extracting data from one application, transforming the data into a format suitable for analysis, and moving it into a data warehouse. There is a wide array of different types of applications that are needed to keep an enterprise functioning.

These different types of applications have different requirements and constraints that must be considered when moving them to the cloud. For example, they might need:

1. To start and finish executing within a particular time period.
2. To wait for another job to complete before it can begin.
3. To limit the functionality of some services, for example, write locking a file to perform a backup.
4. To provision a significant number of servers for a short period of time for a compute intensive operation.

Every Business will have finite resources. As part of the planning process, we need to understand the true cost of delivering a set of workloads. Often, when people compare the cost of deployment options, they limit the comparison to the cost of hardware acquisition. This can be quite misleading. It is important to consider all of the key elements of cost. Even for a Total Cost of Acquisition (TCA) calculation alone, the different elements of cost include software acquisition costs, software S&S (support & subscription) costs, hardware maintenance costs etc., in addition to the base hardware acquisition costs. A Total Cost of Ownership (TCO) calculation typically is much broader and includes many other relevant elements of cost— some being administration/labor costs, systems management software costs, power and cooling costs, facilities costs, refresh costs etc. Another important aspect to consider is the amount of work being done on the two systems. If two systems being compared have been sized upfront so as to guarantee the same amount of work is being done on both, a direct total cost

comparison is valid. In scenarios where we compare systems delivering different amounts of work, it is imperative that we reduce it to a cost per workload comparison to understand true value.

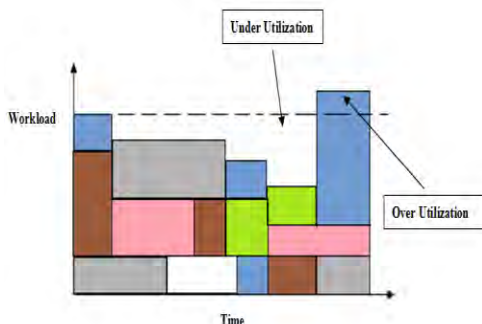


Figure 5. Combinations of workload determining the Total Utilization at a point of time.

The Total Utilization can be increased by minimizing under and Over Utilization. To do so, we can look at existing workloads and try to answer following questions:

1. How often jobs execute on dedicated servers.
2. The level of utilization of those servers.
3. Time constraints on when those jobs execute.

Answering these questions is not an easy task as problem of load balancing arises. For this we want to deploy sufficient cloud infrastructure to avoid periods when cloud consumers want to run more jobs than there is capacity for (over- utilization) at the same time we do not want extended periods of time when servers are idle (under- utilization). This brings us to the second aspect of business alignment: value metrics [4].

B. Value Metrics

The last few years have seen a tremendous increase in server hardware capabilities, especially in the number of processor cores and hardware threads available per server. For example, IBM® POWER7® can support 4 simultaneous threads per core and can scale to 256 cores, making available a massive 1024 hardware threads in a single frame for meeting the ever increasing demand for processing. However, having this kind of hardware capability is not as useful if the software running on that machine cannot exploit those hardware capabilities. The IBM Software suite of products is designed to take better advantage of available hardware computing capabilities. Together with IBM hardware and their differentiated features, IBM Software exploits the higher number of cores and threads, increasing application utilization efficiency, and delivers better price/performance for many workloads.

The figure 6 shows that that the IBM solution is able to deliver an online banking workload at a much lower cost per workload than a competitive option. Thus adding optimization systems with specialized, purpose-built appliances such IBM WebSphere DataPower® for SOA, Netezza for data warehousing, and Cloudburst™ for simplified cloud administration for running specific types of workloads provide focused capabilities with very quick return on investment.

IBM WebSphere Application Server Designed to Leverage Threads in POWER 7 Systems		
WebSphere Application Server v7.0 1 Instance AIX v6.1 64 bit	IBM BladeCenter PS701 8 Cores/3.0 GHz 32 Threads	5,009 Transactions per Second \$32 per transaction per second
Competitive Application Server 1 Instance Solaris 10 64 bit	Oracle Sparc T3-1B 8 Cores/1.65 GHz 64 Threads	746 transactions per second \$137 per transaction per second

Figure 6. How IBM Web Sphere Application Server exploits the hardware capabilities of an IBM POWER7 server.

These appliances are generally comprised of a base operating system, the necessary middleware, and the application in question into a single stack, pre-configured to work out-of-the-box. They typically require lesser skills and labor to configure and operate. Install time is often mentioned in hours instead of days or weeks. Focused delivery capabilities make appliances ideal for certain workload types. Integrating hardware and software to create optimized systems, lowers cost per workload dramatically [5].

C. Advanced Data Capabilities

We are seeing an explosion in data – the volume of information being generated has increased exponentially in the last few years. This data needs to be stored, managed and used efficiently. Storage systems need to be able to handle this growth. Software for managing data needs to leverage hardware capabilities. Business analytics and deriving intelligence from information is becoming an increasingly important factor that decides how competitive a company will be.

IBM software, servers and storage systems are meeting this challenge. IBM offers a broad portfolio of storage systems, ranging from small to mid-range to enterprise class. There are numerous innovations in this space. For example, IBM's Easy Tier® capability can automatically allocate optimum amounts of Solid State Drives (SSD), thus maximizing SSD performance gains while minimizing costs. Another example of innovation in this space is the IBM Smart Analytics Optimizer—a purpose-built appliance for optimizing business analytics on zEnterprise.

Real world Business Analytics workloads need to support concurrent performance. A typical Business Analytics solution will need to support multiple users executing a wide variety of queries and reports concurrently. IBM offers packaged solutions like the IBM Smart Analytics System (ISAS) that have been built as pre-packaged, pre-configured solutions to support this kind of real world usage patterns

Solid state disk drives are revolutionizing storage performance—they support much higher Input/Output Operations (IOPS) than traditional hard disk drives (HDDs). However, SSDs are also more expensive than HDDs. Over provisioning storage systems with SSDs in the hope of getting the best performance often results in huge storage cost increases. We find that, often a small fraction of SSDs will yield the majority of the performance gain possible for a given workload. The ideal price/performance point is reached by having a judicious mix of SSDs and HDDs. Doing this manually is quite

cumbersome and inefficient. IBM's Easy Tier can optimize the amount of SSD allocated. It dynamically moves data to SSD, based on hot spots detected. Further, Easy Tier can dynamically share the available SSDs across many workloads, efficiently allocating the SSDs to the hottest spots.

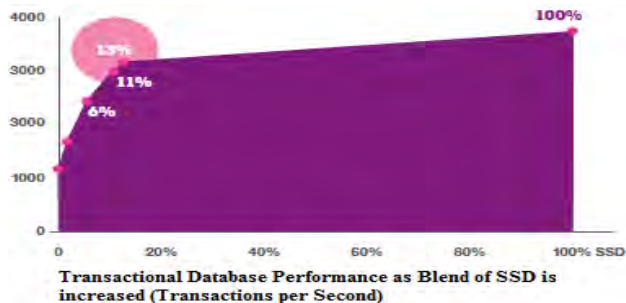


Figure 7. Easy Tier in IBM Storage Systems Automates Optimum Use of SSD

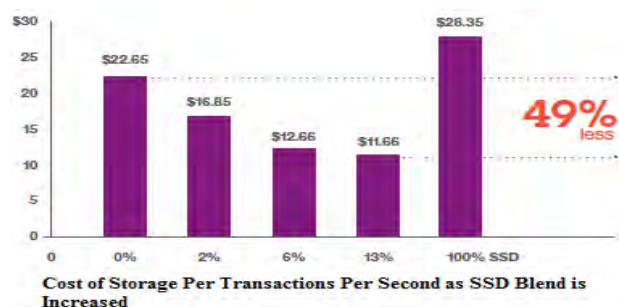


Figure 8. Easy Tier in IBM Storage Systems Automates Optimum Use of SSD (Continued)

Figure 7 & 8 demonstrates how Easy Tier helps improve IT economics when dealing with data workloads. Figure 7 shows just 13% blend of SSD to HDD achieves 171% performance gain and Figure 8 shows Easy Tier achieves 78% of the maximum SSD performance potential with a blend of just 13% SSD.

Automatic allocation of expensive SSD resources in an optimum way results in lowering overall storage costs. Automation also reduces labor costs. IBM offers a comprehensive suite of software products aimed at efficiently automating various data center tasks. Some examples are software that automatically provisions systems, software that synchronizes the start and stop of sequenced applications, software that automatically adjusts system resources available to workloads to meet varying demand etc. Automation not just reduces labor costs; it also improves the quality of service delivery. This is especially true as automating mundane repetitive tasks eliminates the risk of manual operations error [5].

D. Improve IT Economics with Private Clouds and New Service Delivery Models

Cloud computing is poised to transform the service delivery model. At an abstract level, it is about delivering hosted services with some key distinguishing attributes:

1. *Elastic Scaling:* Resources scale up and down by large factors as the demand changes.

2. *Flexible Pricing:* Utility pricing, variable payments, pay-by-consumption and subscription models make pricing of IT services more flexible.
3. *Ease of Use:* End user often just needs a PC with Internet access to request services, without IT skills or knowledge of the system.

Many businesses are moving to a Cloud Computing model. Line-of-business units within organizations are going to public cloud providers as their low cost, pay-per-use model is seen as more cost effective. Furthermore, faster provisioning of resources offered by them can enable users to respond with agility. In some cases, the public cloud model makes the most economic sense [5].

There are three broad modes of delivery for cloud services: private, public, and hybrid. A private cloud is deployed and managed by an organization for its own internal use. The organization controls all aspects of cloud implementation, management, and governance. One of the most significant advantages of this approach is that data never leaves the control of its owner. This reduces the risk that an outside party will gain access to private or confidential data. Depending on the implementation and management details, private clouds may be more cost effective as well. For example, a business may have significant investment in servers that can be redeployed in the cloud, lowering the initial costs.

A public cloud is one that is managed by a third party that provides services to its customers. The primary advantage is low start-up costs on the part of customers and minimal management overhead, at least with respect to basic cloud services. Businesses will still need to manage their workloads, allocate chargeback, and so on. Choosing between public and private cloud implementations is not an all-or-nothing proposition. Hybrid clouds, or the combination of private and public implementations to run business services, have emerged as a third alternative. Consider the economic benefits. There may be a point, however, at which the benefit of adding servers to a private cloud is not sufficient to offset the costs of adding them. For example, the distribution of workloads may entail a number of peak periods where demand exceeds the capacity of the private cloud. These peaks may be regular short periods (for example, at the end of the month when accounts are closed and data warehouses and data marts are updated and many reports are generated) or they may be more unpredictable periods of high demand.

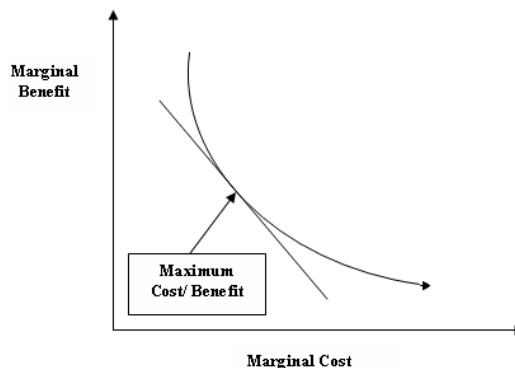


Figure 9. Cost/Benefit of additional Investment in a Private Cloud.

The figure demonstrates the cost of adding and maintaining additional cloud resources eventually reaches a point where the costs outweigh the benefits. At this point, a hybrid cloud approach may be the most cost effective option.

In general, private clouds are a model for delivering IT services in a pay-as-you-go fashion similar to what public clouds can offer. The difference is that a private cloud is built from resources inside an organization and is typically hosted within the data center to be used by line of business and other end users within the organization. Workloads run on large scale virtualization platforms. This approach may significantly reduce hardware, software and labor costs. Users request services via a self-service portal and virtual machines are quickly provisioned. Labor costs may be significantly reduced via standardization and automation. Furthermore, users may be given tools to manage their applications on the virtual machines that are running their services and they only pay for what they consume [4].

For example, IBM provides large scale virtualization environments in POWER7 and zEnterprise that are ideal to host private clouds. IBM provides software that manages the entire lifecycle of virtual servers—everything from self-service automated provisioning to metering and billing based on usage. IBM also offers industry specific Cloud Service Platforms. An example is the recently announced IBM Cloud Service Provider Platform, a comprehensive set of hardware, software and services to help providers rapidly deliver cloud computing on their own.

The zEnterprise system provides the broadest architectural support for building a private cloud. Different environments in zEnterprise may be used to run workloads without requiring a port or rewrite. Where there is an option, workloads may be best fit to an environment to run at the lowest cost per workload. A Fit-for-Purpose deployment strategy aims to assign a workload to the environment that best satisfies the particular requirements of that workload. For example, workloads with heavy IO demand may be best fit on Linux on z/VM on the z196 portion of zEnterprise. Workloads that have high CPU demand and that can exploit multithreading may be best fit on the POWER7 blades in the zBX. Large scale virtualization on z/VM drive down acquisition costs. The private data network as well as the private management network between the z196 and the zBX plus network access control mechanisms ensure a secure network environment. Managing this environment with zManager and with Integrated Service Management (ISM) software, resulting in reduced operational costs [5].

I. CONCLUSION

We started with a question: How can IT better serve the business? There is no one simple answer to this question but it is clear that aligning development and operations processes driven by today's business and technological maturation has exposed new challenges. There are three main challenges when it comes to implementing a Business Intelligence solution: staying within budget, meeting the needs of the business, and efficiently maintaining the solution going forward.

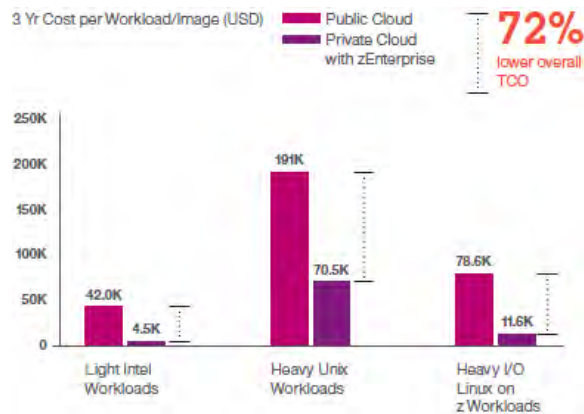


Figure 10. Private Cloud on Z Enterprise Dramatically Reduces Costs using virtualization servers

However, the approach discussed above to reduce the economics of IT and thereby improving the business has produced cost-effective solutions to maintain enterprise-wide workloads, integrate within the existing IT environment, and adapt to changing business needs. The graph below shows that if Business Intelligence is implemented in a Smarter Computing Environment can minimize costs and complexity by leveraging existing IT investments and increase IT productivity by reducing Application Development time, Total cost per workload, Server acquisition cost, Database cost, Power Consumption, New Application Development Time, Floor Space and thereby maximizing organizational effectiveness and helping the Business to grow in a Smarter way as the planet is becoming Smarter.

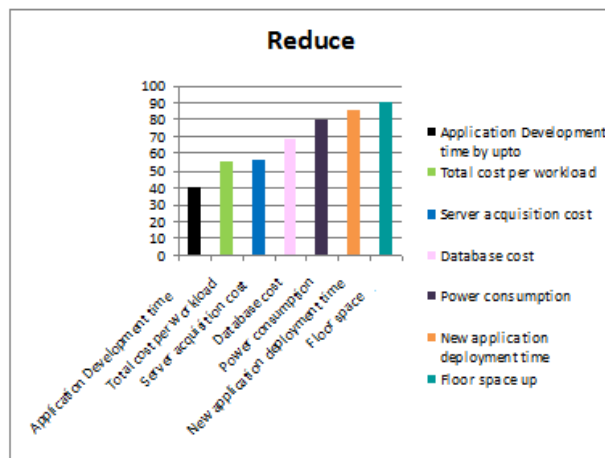


Figure 11. Smarter Computing improving Business efficiency.

II. FUTURE SCOPE

The system accuracy can be improved by adding Data Governance process in place in order to maintain data integrity and the quality of data being used across the enterprise, i.e., being able to control where the data comes from, who controls data values, consistent data types, etc. Having a tight process in place can also help ensure that business data is not exposed to unauthorized personnel, especially sensitive information that can be exploited.

In the past, database administrators had exclusive access to the data and would provide canned reports to the business users. Now business users are being given greater access to data so that they can slice and dice the information and organize the data in a way that best meets their needs. Making your Data Governance processes more robust can not only allow for broader usage of your data to improve insight and decision making capabilities, but it will also help ensure that data quality is maintained in a safe and secure way and using Predictive Analytics to find patterns in historical and transactional data to identify opportunities and risks for the business has become increasingly sophisticated over the last few years [10].

REFERENCES

- [1] IBM, "Deliver smarter products and services by unifying software development and IT operations," Issue Date : Sep. 2009, ISBN RAW14175-USEN-00.
- [2] Hien Nguyen Van, Frederic Dang Tran, "Autonomic virtual resource management for service hosting platforms," Issue Date: May 23, 2009, ISBN 978-1-4244-3713-9.
- [3] Derrick Kondo, Bahman Javadi, Paul Malecot, Franck Cappello, David P. Anderson, "Cost Benefit Analysis of Cloud Computing versus Desktop Grids," Issue Date : Sep. 2009, ISBN 978-1-4244-3750-4.
- [4] Dan Sullivan, "The Definitive Guide to Cloud Computing," Realtime Publishers.
- [5] IBM, "Smarter Computing & Breakthrough IT Economics," Issue Date : Apr. 2011, ISBN XBL03007-USEN-01.
- [6] Gary Barnett, "Mainframe and the Cloud," Issue Date : Nov. 2010, ISBN BC-INF-W-00068UK-EN-00.
- [7] Bola Rotibi, Ian Murphy, "Business Benefits of an Architectural Strategy for Smarter Computing," Research Analyst in Creative Intellect Consulting, Issue Date : Jun. 2011.
- [8] Sam Palmisano, "Smarter Computing -The Next Era of IT," IBM, Issue Date: April 2011.
- [9] G. Khanna, K. Beaty, G. Kar and A. Kochut, "Application Performance Management in Virtualized Server Environments," Network Operations and Management Symposium. 2006.
- [10] Raj Gill, "Top Ten Data Management Trends," Scalability Experts, Issue Date: June 2011.
- [11] "www-03.ibm.com/systems/data/flash/ae/smartercomputing".

AUTHORS PROFILE

Saurabh Kacker received his bachelor's degree in Computer Science from Rajasthan Technical University, Kota, India and currently pursuing Master's Degree from Lingaya's University, Faridabad, India. His areas of interests include Cloud Computing, Network Security, Natural Language Processing etc.

Vandana Choudhary received her bachelor's degree in Computer Science from Kurukshetra University, Kurukshetra, India and currently pursuing Master's Degree from Lingaya's University, Faridabad, India. Her areas of interests include Cloud Computing, Natural Language Processing etc.

Tanupriya Choudhury received his bachelor's degree in CSE from West Bengal University of Technology, Kolkata, India, master's Degree in CSE from Dr. M.G.R University, Chennai, India and currently pursuing his Doctoral Degree. He has two year experience in teaching. Currently he is working as Asst. Professor in dept. of CSE at Lingaya's University, Faridabad, India. His areas of interests include Cloud Computing, Network Security, Data mining and Warehousing, Image processing etc.

Vasudha Vashisht received her bachelor's and master's degree in Computer Science from M.D. University, Haryana, India. She has 6 years of experience in teaching. Currently, she is working as Assistant Professor in the Dept. of Computer Sc. & Engg. at Lingaya's University, Faridabad, Haryana, India. She has authored 10 papers and her areas of interests include artificial intelligence, Cognitive Science, Brain Computer Interface, Image & Signal Processing. Currently she is pursuing her doctoral degree in Computer Science & Engg. She is a member of reputed bodies like IEEE, International Association of Engineers, International Neural Network Society, etc.

A Flexible Tool for Web Service Selection in Service Oriented Architecture

A Flexible Tool for Web Service Selection

Walaa Nagy, Hoda M. O. Mokhtar, Ali El-Bastawissy
Faculty of Computers and Information
Cairo University
Cairo, Egypt

Abstract—Web Services are emerging technologies that enable application to application communication and reuse of services over Web. Semantic Web improves the quality of existing tasks, including Web services discovery, invocation, composition, monitoring, and recovery through describing Web services capabilities and content in a computer interpretable language. To provide most of the requested Web services, a Web service matchmaker is usually required. Web service matchmaking is the process of finding an appropriate provider for a requester through a middle agent. To provide the right service for the right user request, Quality of service (QoS)-based Web service selection is widely used. Employing QoS in Web service selection helps to satisfy user requirements through discovering the best service(s) in terms of the required QoS. Inspired by the mode of the Internet Web search engine, like Yahoo, Google, in this paper we provide a QoS-based service selection algorithm that is able to identify the best candidate semantic Web service(s) given the description of the requested service(s) and QoS criteria of user requirements. In addition, our proposed approach proposes a ranking method for those services. We also show how we employ data warehousing techniques to model the service selection problem.

The proposed algorithm integrates traditional match making mechanism with data warehousing techniques. This integration of methodologies enables us to employ the historical preference of the user to provide better selection in future searches. The main result of the paper is a generic framework that is implemented to demonstrate the feasibility of the proposed algorithm for QoS-based Web application. Our presented experimental results show that the algorithm indeed performs well and increases the system reliability.

Keywords—*Semantic Web; Web services; Web services match-making; Data warehouses; Quality of Services (QoS); Web service ranking.*

I. INTRODUCTION

A. Background

Web services are considered as self-contained, self-describing, modular applications that can be published, located, and invoked across the Web [1]. With the

development of service-oriented architecture (SOA) and cloud computing, more and more services are continuously

emerging on the Internet, such as Amazon EC2¹, Google App Engine²; thus, it is expected that in the near future there would be more and more different types of services, and more and more number of services emerging on the Internet.

Besides as the users often do not know how to quantify the trade-offs between different Web services and just wish to quickly grasp what can be potentially interesting, a single solution that is the best one from an objective point of view typically does not exist; instead, many reasonable alternative services usually exist .

Hence, as both the user requirements, and the number of available services and service providers increases, improving the effectiveness and accuracy of Web service discovery and selection mechanisms becomes a crucial issue [2, 3].

Today, the Universal Description, Discovery and Integration UDDI standard is considered the most commonly used service discovery standard [4]. However, UDDI has 2 main shortcomings: first, it returns coarse results for a keyword based search, and second, more importantly it lacks semantics. Hence, UDDI is basically a framework that supports category based search [4].

On the other hand, semantic Web improves the quality of existing tasks, including Web services discovery, invocation, composition, monitoring, and recovery by describing Web services capabilities and content in a computer interpretable language [4].

One of the main applications of semantic Web is its usage in the semantic Web services in the matchmaking process. Matchmaking is the process of finding an appropriate provider for a requester through a middle agent. Consequently, Semantic matchmaking is used by semantic Web services to find valuable service candidates and selecting the most suitable service(s) that best match user request [5, 6]. In this work we use OWL-S [7] for describing the used services.

OWL-S is an OWL-based Web service ontology, which supplies a core set of markup language, constructs for describing the properties and capabilities of Web services in

¹<http://aws.amazon.com/>

²<http://code.google.com/appengine/>

an unambiguous, computer-interpretable form. The overall ontology consists of three main components: *the service profile* for advertising and discovering services; *the process model*, which gives a detailed description of a service's operation; and *the grounding*, which provides details on how to interoperate with a service, via messages. Specifically, it specifies the signature that is composed of the inputs required by the service, and the outputs generated. Furthermore, since a service may require external conditions to be satisfied, and its execution can change those conditions, the profile describes the preconditions required by the service and the expected effects that result from the execution of the service. For more details, we refer the reader to for example [7].

Nevertheless, with the increasing number of Web services providing similar functionalities, the QoS (Quality of Service) is becoming an important criterion of selection of the best available service. Although, we believe that designing intuitive, easy-to-use user interfaces can help the process of collecting user feedback and preferences; in this work, we do not deal with this issue, instead our focus is on how the collected information in the user profile is processed and integrated in the selection process of Web services to improve the results of subsequent searches.

Inspired by the fact that current enterprise decision making systems benefit more from OLAP, and data warehouse techniques [8], in this paper we show how we can adopt the power of data warehousing in supporting decision making, and how data warehouses and OLAP techniques can help in selecting the most interesting result with the above issues being considered.

In general, data warehousing is one of the most common business intelligence tools nowadays. Data warehouses provide a solid platform that includes both current and historical data [8]. Using this platform, companies can therefore make a series of analysis that can help in providing the right service that matches the user request more easily, accurately, and efficiency.

B. Motivation

The availability of service providers with different features makes the task of selecting an appropriate service provider for a user more and more complex which motivates us to consider new solutions for the Web services selection problem in SOA systems. As shown in Table I there are various types and number of services, associated with different performances, prices, platform/APIs, and availability levels [9, 10, 11].

Analyzing these services we find that:

- 1) *There exist various types of services (for example compute, storage etc).*
- 2) *There exist a large number of functionally similar services which results in a proliferation in the number of services that provide similar functionality with different QoS criteria (i.e. price, availability).*

3) *There are a large number of service providers that are continuously emerging on the internet such as Google AppEngine.*

4) *Finally, there is a wide range in service performance and price. Where different providers offer their services with different prices and performance values.*

From that we can conclude that Web service selection process needs five crucial issues:

a) **Accuracy:** *the algorithm should avoid the loss of Web services that can match the user request but their interface is not the same as the user request. Thus, semantic matchmaking of Web services is needed.*

b) **Flexibility:** *new evolving mechanisms should be flexible to support large numbers of services providers.*

c) **Scalability:** *selection algorithm of Web services should be scalable to support any number of QoS requirements.*

d) **Generality:** *the selection algorithm should be as generic as possible to support different users and various user requirements, rather than specific types of users.*

e) **User personalization:** *the algorithm should be able to provide the right service to the right user request; ideally the user preferences should be captured automatically*

C. Our Contribution

Inspired by importance of the Web service selection problem and its vital role in satisfying the requests of billions internet users, in this paper, we address the service selection problem. We focus on the five challenges that we presented earlier namely, the accuracy, flexibility, scalability, generality, and personalization. Our main contributions are as follows:

1) *We propose a new service selection algorithm that uses a semantic matchmaker to enhance the selection accuracy.*

2) *We include QoS criteria in our selection process to find the service that best matches the user requirements and constraints.*

3) *We employ data warehousing techniques to capture the historical user profile to provide a better service personalization based on previous user requirements and selections.*

4) *We experimentally show that the proposed algorithm enhances the quality and efficiency of the selection process.*

The paper is organized as follows: Section II presents an overview of previous work. Section III describes QoS properties that will be used. Section IV discusses the system architecture. Our proposed selection methodology of Web services is presented in section V. Experimental evaluation is provided in section VI. Finally, section VII concludes our work and presents directions for future work.

Service Provider	Service Type	Price	Platform/API	Availability(SLA)%
Google AppEngine	Compute	\$8/application	Java/Spring/Python	99.9
Azure compute (small)	Compute	0.12/h	Windows server2008	99.95
IBM cloud (Unres. Bronze)	Compute	\$0.210/ h	RedHat Linux	99.5
AWS SimpleDB	Database	\$0.250/GB/Month		
Azure storage	Storage	\$0.15 /GB/Month		99.9

TABLE I. SERVICES PROVIDED BY REPRESENTATIVE PROVIDER

II. RELATED WORK

Today we are witnessing a proliferation in the number of available Web services; this proliferation increases the need for automatic Web service retrieval algorithms. Currently, Web service discovery is a challenging task specially when finding the services that match users' interest. This challenge is a natural consequence of the inability of service

discovery processes to resolve ambiguities introduced by Web service interfaces. Unfortunately, many of the existing discovery models restrict themselves to finding Web services solely based on the descriptions available within WSDL documents [12, 13].

Several approaches have been proposed in the literature for discovering Web services. In [14], the authors proposed a system that discovers Web services based on keyword matching by taking advantage of the IR technique utilizing Vector Space Model (VSM). This approach computes the similarity between query terms and the document collection focusing mainly on WSDL operations (e.g. operation names).

In a similar effort, the authors in [15] proposed Woogle, a search engine which focuses on retrieving WSDL operations retrieving WSDL operations. Woogle (which discontinued its services in 2006), collected services from accessible service registries and provided clients with capabilities to perform keyword-based search. However, the main underlying concept behind the method implemented in Woogle was based on the assumption that Web services belong to the same domain of interest and are equal in terms of their behavior in accomplishing the required functionality.

In [16], the author provided a comprehensive list of QoS parameters that cover the quality in Web services, and classified them into categories including: (1) runtime- QoS attributes, such as scalability, capacity, performance, reliability, availability, robustness, accuracy, and exception handling; (2) transactional- QoS which mainly focuses on the quality of transactions executed (integrity); (3) configuration management and cost-QoS properties that are to standards and cost, such as regulatory, supported standards, stability, cost, and completeness; and (4) security- QoS properties that are to security, such as authentication, confidentiality, accountability, data encryption, traceability, and non-repudiation.

Other researchers have provided similar lists of QoS properties [17, 18, 19], however, little or no details are given on how to calculate or compute the proposed QoS parameters. Recently, a number of approaches were proposed that presented experimental frameworks that attempt to provide QoS measurements and support for Web services. One of the most common frameworks is QoS Certifier introduced in Ran [16] in which a system is proposed for adding QoS information in UDDI registries using a QoS certification framework. The QoS Certifier verifies QoS claims provided by a service provider. Although the proposed solution may provide QoS support for Web service discovery, it has several limitations such as the redundancy of performing QoS measurements which first have to be supplied by the service provider at the time of registration, and then those QoS measurements will eventually be performed by a certification authority. In addition, this solution proposed a major change to the UDDI specification [20] which is problematic at this stage.

In [21], the authors used the concept of classes in their proposed approach named WS-QoS. WS-QoS attempts to address issues, such as service selection and monitoring of QoS for Web services. WS-QoS not only defines several QoS parameters, but also includes network-level QoS parameters such as packet-loss, and network delay [21]. However, in real world, it is likely that clients would be more interested to know the overall QoS of a Web service, and not network-level details.

In [22], the authors proposed QoS support for service-oriented middleware (SOM). In this model, the middle-ware monitors QoS metrics for Web services automatically, and four QoS properties were identified: time, cost, reliability, and fidelity. Similarly, in [23], the authors proposed a model for identifying services based on QoS guarantees. However, in both of these proposed solutions, the authors did not provide an actual implementation of the proposed systems or how QoS metrics are conducted.

Other approaches focused on the semantic support for Web services as presented in [24], the authors proposed a novel approach to integrate services considering only their availability, the functionalities they provide, and their non-functional QoS properties rather than considering the users direct requests. In [25], the authors proposed a solution for this problem and introduced the Web Service Relevancy Function

(WSRF) that is used for measuring the relevancy ranking of a particular Web service based on QoS metrics and client preferences. However one of the challenges in this work is the clients ability to control the discovery process across accessible service registries for finding services of interest, yet semantic matching of services has not been considered.

In [26] the authors proposed heuristic algorithms that can be used to find a near-to-optimal solution more efficiently than exact solutions. The authors proposed two models for the QoS-based service composition problem. Despite the significant improvement of these algorithms compared to exact solutions, both algorithms do not scale with respect to an increasing number of Web services and remain out of the real time requirements.

Unfortunately many of the existing solutions do not provide ways for clients to articulate service queries tailored to their needs. In fact, the existing discovery models do not sufficiently consider end-to-end discovery mechanisms that can provide clients with quality Web services. In addition, existing QoS discovery models do not provide ways to conduct QoS measurements in a transparent and fair manner. In addition, users do not know the history of the service as if it is a reliable service or not. As a result, the user has to try to use the service and see if it can actually provide the required information or not.

Inspired by the mode of the Internet Web search engine, like Yahoo, Google, the authors on [9] design the service providers search engine (SPSE) algorithm. Different with the existing works, which directly schedule the jobs to resources, the algorithm does not make any schedule decision for the job, but is an assistant tool for service selection.

Our algorithm most similar to his idea but in our algorithm after we select the available providers that can match the user request we enhance the result by conducting semantic matching of services which provide more accurate results also we provide a new method for the selection and ranking of the results of our algorithm. In this work we present a solution that aims to overcome many of the limitations of the existing solutions and offers a novel quality-driven discovery and ranking of Web services. Unlike many of the existing QoS discovery models which require major changes to be made to existing standards such as UDDI, our model serves as an assistant tool for service selection.

Our proposed model measures service qualities in an independent and transparent manner, and allows clients to control and manage the discovery process based on QoS properties.

III. QUALITY OF SERVICE (QoS)

QoS criteria are used to differentiate the Web services providing the same functionality during the service selection process. As a user request can be answered by multiple functionally similar services with different level of QoS. One or more non-functional properties can be associated to a Web service. In this work we use the generic QoS criteria as the basis for further discussions which are also used in [9, 27, 28]. In the rest of this section we explore those additional QoS measures in more details.

- **Trust degree:** Trust degree is a kind of a social attribute of the service provider, and implies its reliability or availability level. The authors in [9] proposed to compute the trust degree of each service provider by aggregating several factors, such as, success rate, user's evaluation to the service, availability level in Service Level Agreement (SLA). Calibrating them into a decimal value between 0 and 1, and denoting by TD_i the i th correlative factor. the trust degree of a service provider can be calculated as:

$$Trust = \sum_{i=1}^d w_i \times TD_i \quad (1)$$

Where w_i is the weight for the corresponding factor, and d is the number of factors. We assume that trust degree is used as a decimal score, with a greater value representing more reliable provider.

- **Execution Time:** is the time interval between the time a service request arrives, and the time the corresponding response is generated. The execution time can be estimated by using existing performance estimation techniques, such as history data [29].

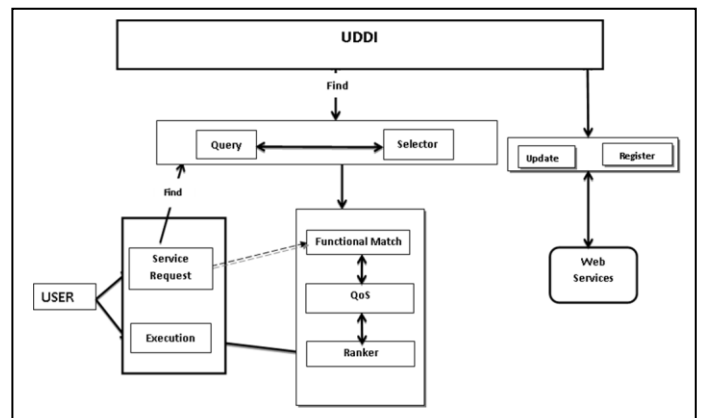


Figure 1. Web Service Selection Architecture

We assume that the response time can be predicted exactly, and simply make use of history data for a service provider's computing power; in case of first time in using services we assume that processing speed standing for provider's computing power.

- **Service Charge:** is the cost that the requester has to pay. The Web service cost can be estimated by operation or by volume of data. i.e. the monetary cost for request execution is commonly defined as:

$$C = D_{in} . P_{in} + T_{exe} . P_{exe} + D_{out} . P_{out} \quad (2)$$

Where, C indicates the total monetary cost, D_{in} represents the data volume transferred into the service provider, T_{exe} stands for the Request execution time, and D_{out} denotes the data volume transferred back to the user after request finish. P_{in} , P_{exe} and P_{out} indicate the prices for transferring in data, job execution and transferring out data, respectively.

- **Service Platform/API:** various platforms and APIs are provided for applications. The users may specify the platform or API requirements (e.g. the .Net applications, or Java/Spring based applications). If the application is originally developed on .Net platform, transplanting it to the Azure can offer tremendous savings in terms of time.

IV. SYSTEM ARCHITECTURE

The main focus of our approach is to design an intelligent system that has the potential of examining Web service's QoS properties in an open and transparent manner, and enabling clients to select the best available Web service by taking advantage of client QoS preferences, Web service capabilities, and service provider features. This is achieved through the following architecture of the proposed solution as shown on Fig. 1.

Web Services from different firms are stored on database using UDDI registry. The service selection algorithm is geographically separated and deployed on the Internet. It communicates with the database to find service providers.

A. Service request module:

This module uses a Web service ontology language (OWL-S) to communicate with the Query module to search for the Web service according to the functional service demand.

B. Services Response module:

This module presents the ranked set of services to the user also to collect users' appraisals towards the founded services. The collected information is offered for further action. by collecting the user's feedback and passing results to the QoS database for adjusting the service provider's appraisal dynamically according to the user's experience To guarantee the data quality, this module will change QoS attribute of user profile values dynamically after each user selection.

1) Query module:

This module communicates with the UDDI Registry to find all the service providers for user's request, and calculates some QoS values, e.g. response time. And store the result.

2) Selector module:

This module will return the service provider candidates; inside the selector we implement our provider selection algorithm.

3) Functional match module:

This module after receiving the services with QoS information, it filter the returned services by performing semantic match between the returned services and services request using OWL-MX matchmaker filters.

4) QoS module:

This module makes inquiries of QoS information regularly from a UDDI repository to check whether any Web service has added or withdrawn its QoS values. It changes information in the QoS database after the new Web service function has been classified that improves to a great extent the quality of the Web service discovery process. To continuously update services QoS values through UDDI, this module provides

reliable service discovery results. In particular, it removes any outdated or broken links.

5) Ranker module:

This module uses the data collected from other modules to generate a ranked list of services. Inside the ranker we implement service filtering and ranking algorithm; using data warehousing techniques to provide decision about selection and ranking of required services by their QoS attributes as required by user. The detailed evaluation process will be discussed later.

6) Finally, Execution module:

This module is in charge of monitoring the execution state of the request. If the service provider is dead, it will be activated to find another service provider to execute the Request. After execution finished, it is also in charge of collecting the results.

V. METHODOLOGY OF WEB SERVICES SELECTION

In this section we explore our proposed approach the proposed system proceeds as follow:

First, given a user request we search for candidate providers that can support this request and then we need to filter the resulted candidate services set generated from the SPSL (Service Provider Selection Algorithm) to

a) *remove bad provider's, and*

b) *to decrease the search space; this is achieved by checking semantic matching between the candidates and service request.*

Then, we perform functional matching (matching input/output parameters) using OWL-MX matchmaking algorithm [14] on the resulting services descriptions from running the SPSL algorithm.

The output of this step is thus a set of matched services with their QoS parameters. Next, we check user profile to get the weights of each QoS parameters and identify the expected user objective function towards the specified parameters. In case users' profile does not include those data we use his class assuming that each user class should contain an importance level towards QoS parameters. Then, we build a data cube whose dimensions are the QoS parameters of the functionally matched services with the aim of maximizing (or minimizing) their values according to the user's objective function. Building the data cube in our model acts as the ranking method for the services providers.

Consequently, the user is provided with a ranked candidate services list with an OLAP report about each service usage to enable him to make efficient decision in selecting the service that could provide the needed information.

Finally, we ask the user for feedback about the results to enhance future requests. In addition, the algorithm considers the case of 2 equivalent candidates; in this case we employ the user rating of those services to select one of them. In case no rating value is available, we provide the user with both services; and with the help of the resulting OLAP report he can decide which one fits his needs. In the following discussion we present the details of the algorithm.

A. Service Model

We assume in this work that services are of type request-response, i.e., they consist of one atomic activity (operation).

A service is represented as below:

Service tuple = $\langle ID, I, O, provider_id, service_type, interface, processing_speed, price, trust_degree... \&others \rangle$

where *ID* is the service identifier, *I* is the set of service inputs, *O* is the set of service outputs, *provider_id* uniquely identifies the provider; *Service type* indicates which kind of service this resource provides; we can access the service through *interface*.

Price indicates the monetary cost that users have to pay for resource utilization; *Trust_degree* represents the provider's reputation; We reserve the other criteria field to support more criteria.

B. Request Model

A service request is used by the framework to select a service provider for a single task from a set of services.

Request = $\langle ReqID; UserID; Input_Data; Services\ type; Interface \rangle$

Where, *UserID* represents the requesters owner; *ReqID* uniquely identifies the request; *Input_Data* is the dataset that need to submit to the service provider; *service type* indicates which kind of service this request needs; and *interface* defines the platform and API that user prefers.

In our work the user request are divided in two parts the first part will be used to generate candidate providers that match specific services type then both the user input data and the discovered candidates are semantically annotated to perform semantic matching between them.

A service request is used by the framework to discover a set of services. Service request and existing services are both described based on common ontology mainly domain ontology and service ontology i.e. OWL-S [7], which allow service discovery and enable interoperability of the discovered services.

A service request consists of a set of semantic annotations (*ID, I, O*) that describe declaratively the desired service properties. *ID* is the request identifier, *I* is the service input, and *O* is the service output.

C. The Service Provider Selection Algorithm (SPSA):

After we select the set of providers that match the user request type as shown in Fig. 2, the number of candidate services is further reduced by checking semantic matching between the candidates and service request, and by adjusting user profiles as we will show in the next section.

D. Service Filtering and Ranking Algorithm

Given the results returned in Section V-C, we filter those candidate services by first performing semantic matchmaking between them and the requested services as follow:

```
Algorithm:Service Providers' Selection Algorithm
Input: user_id:requester, req_id:uniquely identifies the request ,
       service_type:indicates which kind of service this request needs,
       interface:defines the platform and API that the user prefers
Data: The service model of all services providers
Result: Service providers list (SPL)
if The service provider is available & its service type and platform/API are the same
as what's in the request required then
    Search service providers from inside UDDI registry and return set of matched
    SPL list
end
forall the SPi(service provider) ∈ SPL do
    Calculate:{Price,Trust,Performance}
end
return The SPL List /* SPL contain the matched services with
their provider QoS values */
if No results are returned then
    Let the user reconsider the values in his request and start again
end
```

Figure 2. The Service Provider Selection Algorithm

1) Functional matching.

Semantic Web based approaches have been applied to semantically annotate Web Services to allow automated discovery and ranking, followed by mediation and invocation. Rich semantic descriptions allow Service Providers to model their services in a more expressive way that makes it easier for the Service Consumers to search for the required service using semantic reasoning and querying approaches.

Use of semantic annotation is important for appropriate service discovery and to help user specify problem using free text which is translated to semantic description of the problem.

The users *Input_Data* are semantically annotated and also the candidate's services generated from SPSA. The functional part of a semantic Web service can be described by a quadruple $SWS = (I, O, P, E)$, where *I, O, P, E* are sets of Inputs, Outputs, Preconditions, and Effects, with each parameter semantically annotated by means of an associated ontology "O". Matching a service request *R* with a service offer *S* is based on matching the individual parameters in the two descriptions. The service discovery process consists of checking all returned services from SPSA that semantically match the service request inputs, outputs (IO).

The proposed algorithm uses the OWLS-MX service matchmaker [30] to process service requests and advertisements described in OWL-S, and to compute the pairwise similarities between parameters. In particular, we use this matchmaker because it provides five different matching filters. The first performs a purely logic-based match (M0), characterizing the result as exact, plug-in, subsumes, or subsumed-by. The remaining four perform hybrid match, combining the semantic-based matchmaking with the following measures: loss-of-information (M1), extended Jaccard similarity coefficient (M2), cosine similarity (M3), and Jensen-Shannon information divergence based similarity (M4). For each pair (R, S) of a service request and service advertisement, OWLS-MX applies one of the filters M0 – M4, and calculates a single score denoting the degree of match between R and S [30].

2) Non-functional and personalization.

After functional match is performed, we assume there are multiple sources of information for each service request; this implies that each request can be answered from multiple functionally similar Web services, so we need to decide which Web service provider is of higher quality. Hence, after services have been chosen based on functional parameters, non-functional matching is performed and user's profile has to perform another important task while using the service, it has to supply preferences of users towards the values of the parameters that are transferred to the service:

1. To save the time needed to identify the weight of different quality parameters each time user initializes a service request
2. To ensure providing the right service to the user as he expects.

We use user preference in our model assuming that each user profile contains three parts:

- 1) The first part contains the class of user (i.e. Business, economic, social ... etc.).
- 2) The second part consists of the importance level of the preferred quality of service parameters and helps to choose between discovered services. We assume that these values range from 0 to 1.0. For example, a 0.99 value for price, and 0.44 for availability. Those values will be used as a user weight for QoS parameters. Also if the user considers all properties as important, then the weights are distributed equally. If user considers only certain attributes are important then the weights will be distributed equally between the other remaining attributes and also this part include the user preference for business properties of services like payment method.
- 3) The third part deals with the preferred characteristics of objects or information that a service claims to provide. This part helps to choose between different services and to discard services that can in principle handle the task, but do not provide any desirable objects. This is the user rating of service that can be 0 or 1.0.

In our study QoS constraint represents user's end-to-end QoS requirements. These can be expressed in terms of the user objective function towards the different QoS criteria as follows:

If the parameters values in a range less than 0.5, then we assume that the user wants to minimize the values, and if the values in a range greater than 0.5, then, we assume the user wants to maximize the objective function. Those preferences can be gathered from previous interactions in the form of a long-term profile or can be directly specified by the user in the form of soft constraints.

Definition 1: Given a candidate service set for a request denoted by CSS , and a vector C of QoS constraints on CSS given by: $C = c_1, c_2, \dots, c_m$. Let S be an instantiation of CSS in which a concrete Web service is selected. S is a feasible service selection iff S satisfies all QoS constraints in C .

In case of two identical services (i.e. two similar candidates) which candidate should be chosen is a crucial question. Deciding which one better presents the user's request when they are identical requires us to define an optimal service. Thus, it is important to address the problem of finding plans that consistently choose the highest quality available Web services.

Definition 2: An optimal service selection for a given Web service request R , and a given vector of QoS constraints C is a feasible selection with the maximum non-functional matching value. The non-functional matching value of each service is calculated by:

$$(S_i) \sum_{i=1}^d W_i q_i(S) \quad (3)$$

Where, $NF(S_i)$ is the non-functional matching value, W_i is the weight for the QoS parameter identified by the user, and, $q_i(S)$ is the value of the i^{th} QoS parameter in service ' S '.

3) services ranking.

Next, we provide a ranking method to sort all matched candidate services based on user's preferences towards the criteria (time, cost, trust degree)

$$\text{Rank} = \text{Functional Match value} + \text{non-functional match value.} \quad (4)$$

$$\text{Non-functional match} = \sum_{i=1}^d W_i \times q_i(S)$$

In case the users want to minimize the criteria value we multiple the value of non-functional match by (-1).

According to the research direction described in sections I, II we introduce a multi-dimensional user model in which the set of feasible services of user request are organized in an OLAP fashion, such that:

1. The cube dimensions represent the QoS parameters; We use the vector $Qs = q_1(s), \dots, q_r(s)$ to represent the QoS attributes of service ' S ' which define our data cube dimensions, where the function $q_i(S)$ is the value of the i^{th} quality attribute of ' S '.

In reality, companies managing the service searching engines can deploy special applications themselves to obtain their own experience on QoS of some specific Web services. Alternatively, they can also hire third party companies to do these QoS monitoring tasks for them [31].

2. The measure of each dimension is the value of Rank function.

Objective function towards those dimensions are calculated as we mentioned earlier depending on the weight for each value in the quality vector ' C ' the user identifies, to be used during the selection of services.

Note that, we map each item in the quality vector ' C ' into a single real value between 0 and 1, by comparing it with the minimum and maximum possible values of service candidates (for example, the maximum execution price of ' CSS ' can be normalized by selecting the execution price of the most expensive service in the candidate set) to allow a uniform measurement of the multi-dimensional service qualities independent of their units and ranges.

$$M(i) = \frac{q_i(S) - Q_{max}(i)}{Q_{max}(i) - Q_{min}(i)} \quad (5)$$

Where, $M(i)$ is the normalization value for the i^{th} dimension for example “price”, $q_i(S)$ is the quality value of the i^{th} dimension in Web service component ‘S’ in the candidate set ‘CSS’, $Q_{max}(i)$ is the maximum value of the i^{th} dimension in the candidate set ‘CSS’, and $Q_{min}(i)$ is the minimum value of the i^{th} dimension in the candidate set ‘CSS’.

As shown in Fig. 3 each multi-dimensional entry on the model contains the objects needed to accomplish the non-functional selection and retrieving task of services; such objects are dynamically built at the beginning of the user sessions. In more detail, we exploit the OLAP logic model and use the well-known mechanism of aggregations on levels [32]. This allows us to represent and manage user variables with a very high level of granularity. The matched services are returned to user in an ordered list based on user objective function.

With the help of OLAP infrastructure we provide the user with the description and history of the services to decide which one to choose from. Service consumers might want to take a look at the history of these alternatives to make a better decision. If a service had good performance during the last year it might be more trust-worthy than services which have been recently published. This in turn is important to ensure result reliability that can help not only the user who wants to both save time and perform the right selection, but also large organizations that want to save money and time. Consequently, using data warehouse will benefit us in:

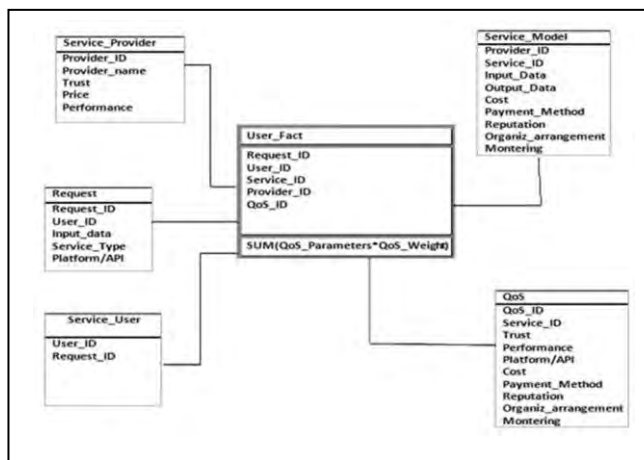


Figure 3. Star schema for non-functional and personalized selection of semantic Web service.

- 1) Past interactions are stored and can be retrieved to save time for search and selection of services.
- 2) Building a data cube with the selected QoS parameters (i.e. execution time, price, etc.), and objective function helps us to select and rank services that best match user request and preference in timely fashion.
- 3) If there are set of matched services that are functionally equivalent with different QoS, the model

will find the best one of them based on user objective function and preference.

- 4) Each cube cell represents constraint over the selected service and so that the objective function to find best matched service can be performed. Also a ranked service list is provided.
- 5) The stored heuristic data about user interaction with different services are extremely useful for service providers as they can display reports about usage percentage of their services during specific periods of time.
- 6) Also, if the system is applied in a company the manager can identify that a specific service provider has poor rating and so does not accept his services in the future.

4) Update users profile.

After we found the service, we enable the user to provide rating on the provided services to show whether these services satisfy his needed information or not. The provided ratings are stored in the system for future use when the same or a similar request is issued.

5) Algorithm Overview.

In this section we explore some basic concepts and notations that we will use in our algorithm. Then, we present our proposed Service Filtering and Ranking algorithm (SFRA) shown on Fig. 4.

Definition 3. A Matched Solution List *MSoL* is a list of service providers that can serve the user request. It is similar, but not identical to the *SPL*:

- a) *MSoL* is generated by filtering the *SPL*, by running *OWL-MX*.
- b) Both the *MSoL* and *SPL* contain the service providers *SP* who will execute user request.

Definition 4. A Final Solution List *FSoL* is a list of service providers generated by filtering the *MSoL* such that: *FSoL* is generated by filtering *MSoL* based on the user profile values. The size of *FSoL* is smaller than the size of *MSoL*.

Definition 5. A Ranked Solution List *RSoL* is a list generated by sorting all the services in *FSoL* such that:

- *RSoL* contains exactly the same services in the *FSoL*.
- Service providers in *RSoL* are sorted using data cube.
- *RSoL* is returned to the end user.

VI. EXPERIMENTAL RESULTS

In this section we discuss our experimental evaluation for our proposed algorithm. In our experiments we manually defined services. Services in the registry are all created and semantically annotated by humans (service developers). We assume in this scenario that all the services are meaningful. All services information’s are stored in a data warehouse which is implemented by files updating instead of a real monitoring service, e.g. MDS service [33], also we used the open source for OWL-MX matchmaker [30] for functional matching of service as described in the functional matching Section.

Algorithm: Services Filtering and Ranking Algorithm (SFRA)
Input: R : User request, SoL Returned from Service Provider Selection Algorithm (SPSA)
 Performs functional service matching using OWL-MX matchmaking filters on SoL to find matched candidate services ($MSoL$) for the specified user request; Groups all matched services with their QoS parameters and description;
forall the QoS parameter p do
 Retrieve the Max and Min value of p from the selected candidate ($MSoL$);
 /* The Max and Min values will be used in normalizing QoS values */
end
if Existent then
 Get the values of the user class and the importance levels (i.e. weight) of QoS parameters from user profile;
end
if Request is found from the past interaction then
 return services matched to this request
end
else
 Calculate the non-functional matching value of each service by:

$$F(ws_i) = \sum (W_j * Q(i, j))$$

 Add Functional and Non-Functional matching values to obtain a single cost function;
 Retrieve $FSoL$ list;
end
if Equivalent services then
 Add user rating value of service to $F(ws_i)$;
 Select Services with maximum $F(ws_i)$;
end
 Build data cube for the $FSoL$ with the aim of maximizing or minimizing QoS values according to user objective function;
return $RSoL$ the ranked results from the data warehouse with an OLAP report about each service usage;
 Let user evaluate the returned results to use this evaluation in the future;
if No results are returned then
 Let the user reconsider the values in his profile and start over again
end

Figure 4. Services Filtering and Ranking Algorithm

In the experiments we simulate different number of service providers ranging from 100 to 13000 providers, we use numeric code indicating the different types of service, uniformly ranging from 1 to 9. The price is evenly generated between 10 and 20 cent per service. And the trust degree is a decimal value uniformly generated from [0; 1].

The experiments were conducted on a DELL Inspiron 1525 machine with 2.80GHz processor Core Duo and 1 GB RAM. The machine is running under Microsoft windows 7 operating system.

In our experiments, we evaluate the performance of our proposed approach through two experiments; the scheduling latency under different number of services providers and different number of QoS criteria.

1. Scheduling latency under different number of services providers: Considering the QoS criteria: response time, monetary cost and trust degree, we evaluate the latency in responding to user request from different number of service providers. Fig. 5, shows that only a small number of service providers are available for the job under our configurations compared to SPSE [9].

Although, in Fig. 5, along with the number of service providers increasing from 500 to 3000, our algorithm has a much higher scheduling efficiency compared to the SPSE

algorithm; our algorithm takes more time for semantic matching of services request as shown in Fig. 6.

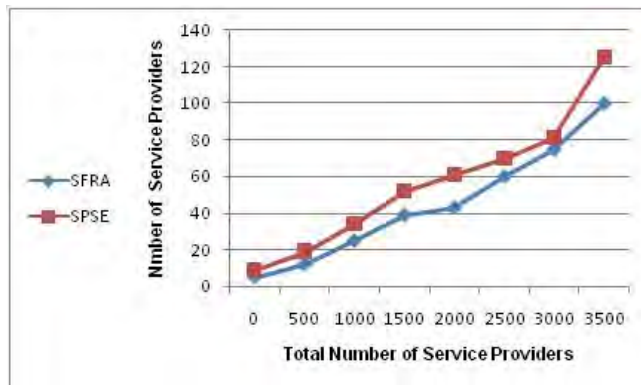


Figure 5. Available service candidates for the request.

For example when we have 2700 service provider our algorithm selects 75 that match the user request in 57:55 sec but SPSE algorithm selects 81 services in 38sec.

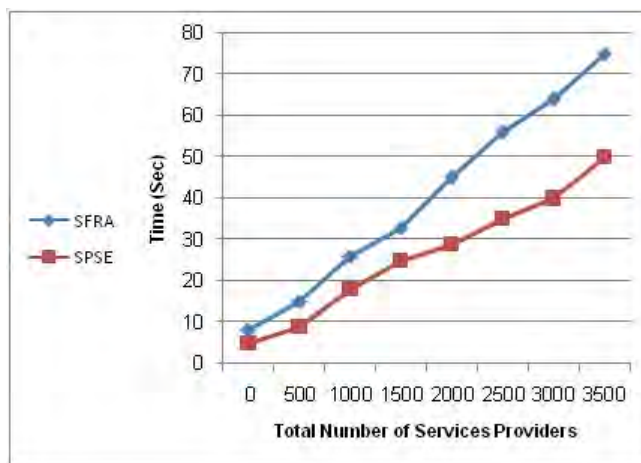


Figure 6. Latency for scheduling 150 request with different number of QoS criteria.

Therefore, our proposed algorithms SFRA is quite useful for user to accurately find the most appropriate service providers but with the cost of increasing the time.

2. The scheduling latency under different number of QoS criteria: Considering 1000 service providers and 150 service requests in the experiments, the SFRA and SPSE algorithms efficiency is also evaluated under different number of QoS criteria. Only the response time is considered in the first experiment, response time and monetary cost are integrated in to the second experiment, and response time, monetary cost and trust degree are implemented in the third experiment. As shown in Fig. 7, the time needed increases linearly with the number of criteria also SPSE algorithm take small time compared to our algorithm.

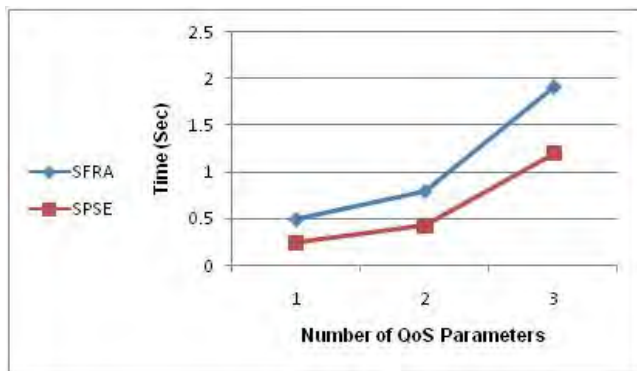


Figure 7. Comparison of SFRA and SPSE algorithms

We summarize the key findings of the comparison test between our approach and the SPSE engine as follows. The first observation is that an optimization of automated Web service discovery techniques appears to be necessary in order to assure the effectiveness of semantic Web services in larger search spaces of available Web services which can be expected in real-world scenarios. The second major outcome is that our approach can be considered as an optimization technique for automated Web Service selection because it can achieve significant improvements in computational performance. Our approach also assures scalability as it supports large number of service providers and QoS parameters, and it demonstrates a high accuracy among several invocations with marginal variations in the number of available services.

VII. CONCLUSIONS AND FUTURE WORK

In this paper we propose an algorithm that is used as a tool in the selection of Web services based on the available providers and user requirements. Our algorithm basically selects the best set of services as required by the user. In addition, the proposed approach ranks those set of candidate services.

A key feature of our proposed approach is that, instead of asking the user for the non-functional properties, the algorithm uses the importance level for QoS parameters already stored in user profile, which makes the algorithm easy to use even for someone not very familiar with the different quality of services attributes.

Besides, the approach allows the user to rate any of the matched services, indicating how relevant or appropriate they are for his request. This rating is used to speed-up similar future searches.

For future work we aim to improve our Web service selection approach and allow it capturing maximum possible and relevant information from publicly available information in user's social network. Social aspects of the information from a network of service consumers and service providers can help a lot in ranking the best available Web services for the users. In addition, we want to consider the case where no matched services are found that match user QoS constraints. We want to explore the effect of relaxing those QoS constraints. Finally, we aim to investigate more on how we

can take into account the users' opinion on the identified objective function to achieve better service matching.

REFERENCES

- [1] J. Rao and X. Su, "A survey of automated Web service composition methods," in Proceedings of the first International Workshop on Semantic Web Services and Web Process Composition, San Diego, California, USA, 2004, pp. 43–54.
- [2] A. Averbakh, D. Krause, and D. Skoutas, "Exploiting user feedback to improve semantic Web service discovery," in ISWC 2009: Proceedings of the 8th International Semantic Web Conference. Springer-Verlag, 2009, pp. 33–48.
- [3] R. Krummenacher, M. Hepp, A. Polleres, C. Bussler, and D. Fensel, "How or what is wrong with Web services discovery," in Proceedings of the Third European Conference on Web Services, ser. ECO '05. Washington, DC, USA: IEEE Computer Society, 2005.
- [4] D. Martin, M. Burstein, G. Denker, J. Hobbs, L. Kagal, O. Lassila, D. McDermott, S. McIlraith, M. Paolucci, B. Parsia, T. Payne, M. Sabou, E. Sirin, M. Solanki, S. Srinivasan, and K. Sycara, "Bringing semantics to Web services: The OWL-S approach," in first International Workshop on Semantic Web Services and Web Process Composition (SWSWPC 2004), San Diego, CA, 2004, pp. 243–277.
- [5] A. B. Bener, O. Volkan, and I. E. Savas, "Semantic matchmaker with precondition and effect matching using SRL," Expert Syst. Appl., vol. 36, no. 5, pp. 9371–9377, 2009.
- [6] L. Cabral, J. Domingue, E. Motta, T. Payne, and F. Hakimpour, "Approaches to semantic Web services: An overview and comparisons," The Semantic Web Research and Applications, vol. 3053, pp. 225–239, 2004.
- [7] "Owl-s: Semantic markup for Web services," 2005.
- [8] K. Pyar, "Decision support system for personnel information using data warehouse," in The 2nd Int. Conference on Computer and Automation Engineering (ICCAE). IEEE, 2010, pp. 668–672.
- [9] L. Zhaoa, Y. Renc, M. Lib, and K. Sakuraia, "Flexible service selection with user-specific QoS support in service-oriented architecture," network and computer application, 2011.
- [10] H. Q. Yu and S. Reiff-Marganiec, "Non-functional property based service selection: A survey and classification of approaches," NonFunctional Properties and Service Level Agreements in Service Oriented Computing Workshop collocated with The 6th IEEE European Conference on Web Services, vol. 411, pp. 13–25, 2008.
- [11] D. A. Menascé, E. Casalicchio, and V. Dubey, "On optimal service selection in service oriented architectures," Performance Evaluation, vol. 67, no. 8, pp. 659–675, 2010.
- [12] S. Agarwal and R. Studer, "Automatic matchmaking of Web services," in Proceedings of the IEEE International Conference on Web Services. IEEE Computer Society, 2006, pp. 45–54.
- [13] H. Q. Yu and S. Reiff-Marganiec, "A backwards composition context based service selection approach for service composition," 2009 IEEE International Conference on Services Computing, pp. 419–426, 2009.
- [14] C. Platzer and S. Dustdar, "A vector space search engine for Web services," in Proceedings of the Third European Conference on Web Services, ser. ECO '05. Washington, DC, USA: IEEE Computer Society, 2005, pp. 62–71.
- [15] X. Dong, A. Halevy, J. Madhavan, E. Ramezani, and J. Zhang, "Similarity search for Web services," in Proceedings of the Thirtieth international conference on Very large data bases, ser. VLDB '04. VLDB Endowment, 2004, pp. 372–383.
- [16] S. Ran, "A model for Web services discovery with QoS," SIGecom Exch., vol. 4, pp. 1–10, March 2003.
- [17] L. Zeng, B. Benatallah, M. Dumas, J. Kalagnanam, and Q. Z. Sheng, "Quality driven Web services composition," in Proceedings of the 12th international conference on World Wide Web, ser. WWW '03. New York, NY, USA: ACM, 2003, pp. 411–421.
- [18] D. A. Menascé, "QoS issues in Web services," IEEE Internet Computing, vol. 6, pp. 72–75, November 2002.

- [19] D. Liu, Z. Shao, C. Yu, and G. an, "A heuristic qos-aware service selection approach to Web service composition," 2009 Eighth IEEEACIS International Conference on Computer and Information Science, pp.1184–1189, 2009.
- [20] "UDDI version 3.0.2 specifications," [http: uddi.xml.org](http://uddi.xml.org) , 20 .
- [21] M. Tian, A. Gramm, T. aumowicz, H. Ritter, and J. Schiller, "A concept for QoS integration in Web services," in Proceedings of the Fourth international conference on Web information systems engineering workshops, ser. ISE '03. ashington, DC, USA: IEEE Computer Society, 2003, pp. 149–155.
- [22] A. Sheth, J. Cardoso, J. Miller, and K. Kochut, "QoS for service-oriented middleware," in Proceedings of the Sixth orld Multi-Conference on Systemics, Cybernetics and Informatics (SCI02), 2002, pp. 528–534.
- [23] Y. Liu, A. H. gu, and L. Z. Zeng, "QoS computation and policing in dynamic Web service selection," in Proceedings of the 3th international World Wide Web conference on Alternate track papers & posters, ser. Alt. '04. ew York, Y, USA: ACM, 2004, pp. 66–73.
- [24] . Ibrahim, . L. Mou'el, and S. r'erot, "Mysim: a spontaneous service integration middleware for pervasive environments," in Proceedings of the 2009 international conference on Pervasive services, ser. ICPS '09. ew York, Y, USA: ACM, 2009, pp. –10.
- [25] E. Al-Masri and Q. H. Mahmoud, "Discovering the best Web service," in Proceedings of the 16th international conference on World Wide Web, ser. '07. ew York, Y, USA: ACM, 2007, pp. 1257–1258.
- [26] T. Yu, Y. Zhang, and K.-J. Lin, "Efficient algorithms for Web services selection with end-to-end QoS constraints," ACM Trans. Web, vol. 1, May 2007.
- [27] L. Zeng, B. Benatallah, A. H. Ngu, M. Dumas, J. Kalagnanam, and H. Chang, "QoS-aware middleware for Web services composition," IEEE Trans. Softw. Eng., vol. 30, pp. 311–327, May 2004.
- [28] D. A. Menasce, "Qos issues in Web services," IEEE Internet Computing, vol. 6, no. 6, pp. 72–75, 2002.
- [29] S. hye Jang, V. Taylor, X. Wu, and M. Prajugo, "Performance predictionbased versus load-based site selection: quantifying the difference," in Proceedings of the 8th international conference on parallel and distributed computing systems, Las Vegas, Nevada, 2005.
- [30] M. Klusch, B. ries, and K. Sycara, "O LS-MX: A hybrid semantic Web service matchmaker for OWL-S services," Web Semantics: Science, Services and Agents on the World Wide Web, vol. 7, no. 2, pp. 121 –133, 2009.
- [31] . Ahmadi and . Binder, " lexible matching and ranking of Web service advertisements," in Proceedings of the 2nd workshop on Middleware for service oriented computing: held at the ACM/IFIP/USENIX International Middleware Conference, ser. M 4SOC '07. ew York, Y, USA: ACM, 2007, pp. 30–35.
- [32] W. H. Inmon, Building the Data Warehouse, 3rd Edition. New York, NY, USA: John Wiley & Sons, Inc., 2002.
- [33] "GT information services monitoring &discovery system MDS," <http://www.globus.org/toolkit/mds/>, 2011.