IJCSI

International Journal of Computer Science Issues

Volume 10, Issue 5, No 1, September 2013
ISSN (Online): 1694-0784
ISSN (Print): 1694-0814

www.IJCSI.org

IJCSI proceedings are currently indexed by:

IJCSI Publicity Board 2013

IJCSI Editorial Board 2013

EDITORIAL

In this fifth edition of 2013, we bring forward issues from various dynamic computer science fields ranging from system performance, computer vision, artificial intelligence, software engineering, multimedia, pattern recognition, information retrieval, databases, security and networking among others.

Considering the growing interest of academics worldwide to publish in IJCSI, we invite universities and institutions to partner with us to further encourage open-access publications.

As always we thank all our reviewers for providing constructive comments on papers sent to them for review. This helps enormously in improving the quality of papers published in this issue.

Google Scholar reported a large amount of cited papers published in IJCSI. We will continue to encourage the readers, authors and reviewers and the computer science scientific community and interested authors to continue citing papers published by the journal.

Apart from availability of the full-texts from the journal website, all published papers are deposited in open-access repositories to make access easier and ensure continuous availability of its proceedings free of charge for all researchers.

We are pleased to present IJCSI Volume 10, Issue 5, No 1, September 2013 (IJCSI Vol. 10, Issue 5, No 1). The acceptance rate for this issue is 30.4%.

IJCSI Editorial Board
September 2013 Issue
ISSN (online): 1694-0784
ISSN (print): 1694-0814

www.IJCSI.org

IJCSI Reviewers Committee 2013

Mr. M. Azath, Anna University, India

Dr. P. Suresh Varma, Adikavi Nannaya University, India

Prof. V. N. Kamalesh, JSS Academy of Technical Education, India

Dr. D Gunaseelan, Ibri College of Technology, Oman

Mr. Sanjay Kumar Anand, CDAC, India

Mr. Akshat Verma, CDAC, India

Mrs. Fazeela Tunnisa, Najran University, Kingdom of Saudi Arabia

Mr. Hasan Asil, Islamic Azad University Tabriz Branch (Azarshahr), Iran

Prof. Dr Sajal Kabiraj, Fr. C Rodrigues Institute of Management Studies (Affiliated to University of Mumbai, India), India

Mr. Syed Fawad Mustafa, GAC Center, Shandong University, China

Dr. Natarajan Meghanathan, Jackson State University, Jackson, MS, USA

Prof. Selvakani Kandeeban, Francis Xavier Engineering College, India

Mr. Tohid Sedghi, Urmia University, Iran

Dr. S. Sasikumar, PSNA College of Engg and Tech, Dindigul, India

Dr. Anupam Shukla, Indian Institute of Information Technology and Management Gwalior, India

Mr. Rahul Kala, Indian Institute of Inforamtion Technology and Management Gwalior, India

Dr. A V Nikolov, National University of Lesotho, Lesotho

Mr. Kamal Sarkar, Department of Computer Science and Engineering, Jadavpur University, India

Prof. Sattar J Aboud, Iraqi Council of Representatives, Iraq-Baghdad

Dr. Prasant Kumar Pattnaik, Department of CSE, KIST, India

Dr. Mohammed Amoon, King Saud University, Saudi Arabia

Dr. Tsvetanka Georgieva, Department of Information Technologies, St. Cyril and St. Methodius University of Veliko Tarnovo, Bulgaria

Mr. Ujjal Marjit, University of Kalyani, West-Bengal, India

Dr. Prasant Kumar Pattnaik, KIST,Bhubaneswar,India, India

Dr. Guezouri Mustapha, Department of Electronics, Faculty of Electrical Engineering, University of Science and Technology (USTO), Oran, Algeria

Mr. Maniyar Shiraz Ahmed, Najran University, Najran, Saudi Arabia

Dr. Sreedhar Reddy, JNTU, SSIETW, Hyderabad, India

Mr. Bala Dhandayuthapani Veerasamy, Mekelle University, Ethiopa

Mr. Arash Habibi Lashkari, University of Malaya (UM), Malaysia

Mr. Rajesh Prasad, LDC Institute of Technical Studies, Allahabad, India

Mr. Sunil Kashibarao Nayak, Bahirji Smarak Mahavidyalaya, Basmathnagar Dist-Hingoli., India

Prof. Nikhil gondaliya, G H Patel College of Engg. & Technology, India

Mr. Nisheeth Joshi, Apaji Institute, Banasthali University, India

Mr. Nizar, National Ingineering School of Monastir, Tunisia

Prof. R. Jagadeesh Kannan, RMK Engineering College, India

Prof. Rakesh.L, Vijetha Institute of Technology, Bangalore, India

Mr B. M. Patil, Indian Institute of Technology, Roorkee, Uttarakhand, India

Dr. Intisar A. M. Al Sayed, Associate prof./College of Science and IT/Al Isra University, Jordan

Mr. Thipendra Pal Singh, Sharda University, K.P. III, Greater Noida, Uttar Pradesh, India

Mrs. Rajalakshmi, JIITU, India

Mr. Shrikant Ardhapurkar, Indian Institute of Information Techonology, India

Ms. Hemalatha R, Osmania University, India

Mr. Hadi Saboohi, University of Malaya - Faculty of Computer Science and Information Technology, Malaysia

Mr. Sunil Kumar Grandhi, Maris Stella College, India

Prof. Shishir K. Shandilya, NRI Institute of Science & Technology, INDIA

Dr. Umesh Kumar Singh, Vikram University, Ujjain, India

Prof. Prasun Ghosal, Bengal Engineering and Science University, India

Dr. Nagarajan Velmurugan, SMVEC/Pondicherry University, India

Dr. R. Baskaran, Anna University, India

Dr. Wichian Sittiprapaporn, Mahasarakham University College of Music, Thailand

Mr. Lai Khin Wee, Universiti Teknologi Malaysia, Malaysia

Dr. Kamaljit I. Lakhtaria, Atmiya Institute of Technology, India

Mrs. Inderpreet Kaur, PTU, Jalandhar, India

Mr. Palaniyappan, K7 Virus Research Laboratory, India

Mr. Guanbo Zheng, University of Houston, main campus, USA

Mr. Arun Kumar Tripathi, Krishna Institute of Engg. and Tech-Ghaziabad, Affilated to UPTU, India

Mr. Iqbaldeep Kaur, PTU / RBIEBT, India

Mr. Amit Choudhary, Maharaja Surajmal Institute, New Delhi, India

Mrs. Vasudha Bahl, Maharaja Agrasen Institute of Technology, Delhi, India

Dr. Ashish Avasthi, Uttar Pradesh Technical University, India

Ms. A. Lavanya, Manipal University, Karnataka, India

Dr. Jalal Laassiri, MIA-Laboratory, Faculty of Sciences Rabat, Morocco

Mr. Ganesan, Sri Venkateswara college of Engineering and Technology, Thiruvallur, India

Mr. V.Ramakrishnan, Sri Venkateswara college of Engineering and Technology, Thiruvallur, India

Prof. Vuda Sreenivasarao, St. Mary's college of Engg & Tech, India

Prof. Ashutosh Kumar Dubey, Assistant Professor, India

Dr. R.Ramesh, Anna University, India

Mr. Ali Khadair HMood, University of Malaya, Malaysia

Dr. Vimal Mishra, U.P. Technical Education, India

Mr. Ranjit Singh, Apeejay Institute of Management, Jalandhar, India

Mrs. D.Suganyadevi, SNR SONS College (Autonomous), India

Mr. Prasad S.Halgaonkar, MIT, Pune University, India

Mr. Vijay Kumar, College of Engg. and Technology, IFTM, Moradabad(U.P), India

Mr. Mehran Parchebafieh, Douran, Iran

Mr. Anand Sharma, MITS, Lakshmangarh, Sikar (Rajasthan), India

Mr. Amit Kumar, Jaypee University of Engineering and Technology, India

Prof. B.L.Shivakumar, SNR Sons College, Coimbatore, India

Mr. Mohammed Imran, JMI, India

Dr. R Bremananth, School of EEE, Information Engineering (Div.), Nanyang Technological University, Singapore

Prof. Vasavi Bande, Computer Science and Engneering, Hyderabad Institute of Technology and Management, India

Dr. S.R.Balasundaram, National Institute of Technology, India

Dr. Prasart Nuangchalerm, Mahasarakham University, Thailand

Dr. M Ayoub Khan, C-DAC, Ministry of Communications & IT., India

Dr. Jagdish Lal Raheja, Central Electronics Engineering Research Institute, India

Mr G. Appasami, Dept. of CSE, Dr. Pauls Engineering College, Anna University - Chennai, India

Mr Vimal Mishra, U.P. Technical Education, Allahabad, India

Mr. Amin Daneshmand Malayeri, Young Researchers Club, Islamic AZAD University, Malayer Branch, Iran

Dr. Arti Arya, PES School of Engineering, Bangalore (under VTU, Belgaum, Karnataka), India

Mr. Pawan Jindal, J.U.E.T. Guna, M.P., India

Dr. Soumen Mukherjee, RCC Institute of Information Technology, India

Dr. Hamid Mcheick, University of Qubec at Chicoutimi, Canada

TABLE OF CONTENTS

1. Multiple Tree Multicast in a Dynamic Environment David A. Johnston, David R. Mcintyre, Francis G. Wolff and Christos A. Papachristou	1-7
2. Detection of Syntactic Aspect Interaction in UML State Diagrams Using Critical Pair Analysis in Graph Transformation Zaid Tahat, Tzilla Elrad, Luay Tahat and Nada Almasri	8-19
3. Simulation Techniques and Control Schemes in Electromyography for a Lower Limb and Power Assist Exoskeleton Ngoc Khoat Nguyen and Thi Mai Phuong Dao	20-24
4. Supporting Context-Centric Relations in Heterogeneous Environments Jamie Walters, Theo Kanter and Rahim Rahmani	25-34
5. Extract an Essential Skeleton of a Character as a Graph from a Character Image Kazuhisa Fujita	35-39
6. Emulating Trust Zone in Android Emulator with Secure Channeling Arun Muthu, Rahim Rahmani and Dinakaran Rajaram	40-51
7. A speech recognition based approach for development in C++ Mubbashir Ayub and Muhammad Asjad Saleem	52-56
8. Handwriting and Hand Drawing Velocity Modeling by Superposing Beta Impulses and Continuous Training Component Houcine Boubaker, Aymen Chaabouni, Najiba Tagougui, Monji Kherallah and Adel M. Alimi	57-63
9. Adapting IS Strategic Planning methodology to define Mobile Strategy Mouhsine Lakhdissi, Rachid El Habi, Alami Semma and Bouchaib Bounabat	64-70
10. Algorithms for Data Compression in Wireless Computing Systems Upasana Mahajan and Prashanth C.S.R	71-77
11. A Generalized symmetric single step method for solving interval linear systems Hassan Badry Mohamed Ahmed El-Owny	78-82
12. Artificial Neural Networks in Medical Images for Diagnosis Heart Valve Diseases Atta Elalfi, Mohamed Eisa and Hosnia Ahmed	83-90
13. Proposed Software Testing Using Intelligent techniques (Intelligent Water Drop (IWD) and Ant Colony Optimization Algorithm (ACO)) Baraa	91-97
14. ERP and Customization: Case Study of Logistics Processes Integration in a Telecommunications Company Mustapha El Masbahi and Alami Semma	98-107
15. Fast Color Edge Detection Algorithm Based on Similarity Relation Matrix Hossam Eldin Ibrahim Ali	108-113

16. In-orbit Calibration Method Based on Empirical Model for Non-collinear TDI CCD Camera **Shujuan Zhang and Shuying Jin**	**114-120**
17. A 3D ear recognition method based on auricle structural feature **Kai Wang and Zhi-Chun Mu**	**121-127**
18. Frequency-Domain Modeling and Simulation of Coupled Lossy Multiconductor Transmission Lines **Youssef Mejdoub, Hicham Rouijaa, Mohamed Saih and Abdelilah Ghammaz**	**128-133**
19. Building web crawler based on bee swarm intelligent algorithm **Muhamed Ibrahim Shujaa and Ahmed Bahaa Ulddin**	**134-141**
20. Cylindrical heat exchanger trajectory planning and tracking using orthogonal functions **Hajer Sayem, Naceur Benhadj Braiek and Hassan Hammmouri**	**142-149**
21. Comparative Study Based on Open Source Content Management Systems Mambo and his Fork - Joomla and Elxis **Dejan Viduka, Igor Lavrnic and Ana Basic**	**150-155**
22. An Efficient and Secure Handover Protocol for IEEE 802.16m Networks **Heba K. Aslan**	**156-164**
23. A Cipher based on Multiple Circular Arrays **Pushpa R. Suri and Sukhvinder Singh Deora**	**165-175**
24. IT-infrastructure of university based on cloud computing **Zhuldyz B. Kalpeyeva and Akkyz K. Mustafina**	**176-179**
25. An Effective Genetic Algorithm for Job Shop Scheduling with Fuzzy Degree of Satisfaction **Akeela M.Al-Atroshi, Sama T. Azez and Baydaa S.Bhnam**	**180-185**
26. Hybrid Genetic Algorithm and Local Search for Energy Demand Prediction Model **Wahab Musa, Ku Ruhana Ku-Mahamud and Azman Yasin**	**186-190**
27. Subjectivity Classification using Machine Learning Techniques for Mining Feature-Opinion Pairs from Web Opinion Sources **Ahmad Kamal**	**191-200**
28. Control Theory based Approach for the Improvement of Integrated Business Process Interoperability **Abderrahim Taoudi, Bouchaib Bounabat and Badr Elmir**	**201-207**
29. Service Orchestration Algorithm for Web Services: Evaluation and Analysis **Yousra Chtouki, Hamid Harroud, Mohammed Khalidi Idrissi and Samir Bennani**	**208-218**
30. Modeling Methodology for NoC : MM4NoC **Sefrioui Nassim, Berrahou Aissam, Diouri Ouafaa and Eleuldj Mohsine**	**219-222**
31. Importance of Software Documentation **Noela Jemutai Kipyegen and William P. K. Korir**	**223-228**

32. A Symbol Based Graphical Schema Resistant to Peeping Attack **T.Srinivasa Ravi Kiran, K.V.Samabasiva Rao, M.Kameswara Rao and A.Srisaila**	**229-233**
33. New Approach to optimize the time of Association Rules Extraction **Thabet Slimani**	**234-243**
34. Adaptive Neuro-Fuzzy Inference System (ANFIS) Based Software Evaluation **Khyati Mewada, Amit Sinhal and Bhupendra Verma**	**244-250**
35. Metric Base Analysis and Modeling Experiments of Routing Protocols in MANETs and VANETs Wireless Network using Real Time Scenarios **Syed Fakhar Abbas, Ghulam Yasin, Muhammad Akram Mujahid and Saqib Rasool Chaudhry**	**251-257**
36. Mapping Wireless Sensor Network Applications Requirements to Existing Operating Systems **Tarek Mohamed Ibrahim Salem and Sherine Mohamed Abd El-Kader**	**258-267**
37. Challenges and Opportunities for the Implementation of Social Network Technologies (SNTs) in Teaching in Universities in Ghana **Adu-Manu Kofi Sarpong, Arthur John Kingsley and Clement Yeboah**	**268-275**
38. BER analysis of 3x3 Spatial Multiplexing under AWGN and Rician channels for different modulation techniques **Anuj Vadhera and Lavish Kansal**	**276-283**
39. Network research based on the fuzzy comprehensive evaluation model of natural language **Zhi-Hong Ma, Xun-Song He and Hao-Xuan Ding**	**284-290**

Multiple Tree Multicast in a Dynamic Environment

David A. Johnston, David R. McIntyre[1], Francis G. Wolff, Christos A. Papachristou

EECS Department, Case Western Reserve University
Cleveland, Ohio, USA 44106

[1] CS Department, Cleveland State University
Cleveland, Ohio, USA 44115

Abstract

Multiple multicast trees have been shown to increase the performance of data distribution when compared with single tree multicast. Node loss and congestion changes the performance characteristics of the multicast trees. Multicast tree performance feedback can be used to determine the optimal tree to use based on the feedback. We further examine an optimizing methodology, Probabilistic Multicast Trees (PMT), for multiple multicast trees which makes use of the performance feedback, generates a probability of usage for each multicast tree based on that feedback and then makes intelligent choices about which multicast tree to use for a given packet in the presence of node loss and congestion.

Keywords: *Dynamic Multicast, Application-Level Multicast, ALM, Probabilistic Multicast Trees, PMT, Adaptive tree selection, Content distribution.*

1. Introduction

Smart phones, movies on demand, regulated industrial process information; the thirst for data access has never been greater and will only continue to grow. Network infrastructure must continue to evolve to meet the ever increasing demand for data. This is especially true when many devices demand the same data at the same time. Since improvement in pure network bandwidth capabilities is only part of the solution; many researchers have investigated efficient transfer of information and a variety of solutions have been proposed. One popular solution is hardware-based methods to distribute this data which resulted in IP multicast. Unfortunately, IP multicast has several limitations that prevent it from being globally used across multiple service provider domains on the Internet. Application level multicast (ALM) overcomes is weakness of being tied down to particular hardware solution [8]. ALM is a multicast overlay network which can be described as a tree. Saltzer [16] argued that the network should be kept as simple as possible and for any multicasting that the intelligence resides at the application layer. ALM is the fundamental principle of the “end to end” argument that Saltzer proposed.

The dynamic behavior of multicast networks presents unique challenges for data distribution in a network environment. Some multicast methodologies repair the multicast trees as needed in the presence of failures. Other multicast systems improve the performance of the multicast tree through probing methods. Both methods aim to address the issue of node loss and congestion. Other multicast research has attempted to address the long delays and performance issues of node loss and congestion by using redundant paths, additional replication of data or using wholly redundant trees. However, in all cases multicast node failures and network congestion still cause long delays, performance issues or missing data as the data is delivered.

In general, *multiple* multicast trees have been shown to benefit multicasting applications in that they increase throughput and reliability [3][10][17].

Many single multicast tree solutions and multiple multicast tree solutions have been developed; however, we still need to make these solutions more efficient. Several approaches to multiple tree multicasting have been implemented [3][4][7][8][15]. These approaches were designed with two goals. The first is to improve performance over the single multicast tree approach and the second is to manage node loss which is a fundamental problem of single multicast trees. Other techniques to manage node loss that were built upon multiple multicasting methods include replication of packets besides just the expected distribution through the tree, forward error correction [14] and multiple description coding (MDC) [11]. All of these schemes, which use additional network bandwidth, address the inherent lossy nature of wireless networks.

We explore Probabilistic Multicast Trees (PMT) [12][13] as applied in a dynamic network environment. PMT is an optimizing mechanism that is intended to improve the capabilities of any multiple multicast tree methodology with respect to the management of node loss and network congestion. PMT is designed to provide two main advantages over other multiple multicast tree schemes. It improves both data delivery latency, and data delivery efficiency.

Data delivery latency, ML_t, is an important performance measure for multimedia streaming. It is the total summation of all the source-to-destination packet delivery times for multicast tree *t*. The time difference can be calculated from a timestamp, T_s, that the source puts into each packet and the receive time, T_r, of the same packet by the receiving destination client. The goal of PMT is to reduce this latency on average over all the receiving clients. Data delivery latency can be expressed by the following equation where the summation is taken over all packets received.

$$ML_t = \sum_r T_r - T_s \qquad (1)$$

Data delivery efficiency (ME_t) refers to the percentage of the total number of multicast packets delivered (P_d) to all client destinations compared to the total number of packets sent (P_s) by the source as expressed by the following equation.

$$ME_t = P_d / P_s \qquad (2)$$

In this paper, we extend our previous work on PMT by taking into account node loss. PMT increases data delivery efficiency by delivering a higher percentage of the packets based on improved multicast tree selection. PMT achieves this by more severely punishing trees with drop out nodes by adding increased feedback delays to these trees resulting in the overall latency feedback for any tree containing such nodes being significantly increased. This results in PMT tending to chose alternate lower latency trees with fewer lost nodes for future transmissions which ultimately is reflected in reduced data delivery latency.

The remainder of this paper is laid out as follows: Section 2 discusses node failure and congestion, Section 3 describes the design of PMT, Section 4 describes data metrics, Section 5 shows the results, and Section 6 discusses conclusions.

2. Node Failure and Congestion Simulation

Past multicasting research has focused on three main areas: building trees efficiently, reducing maintenance overhead, and using other forms besides trees to deliver the data. Multicast overlay network failures causing long delays and performance issues as the data is delivered have been only moderately addressed. Most approaches have either supported repairing the tree as failures occurred or improving the performance of the tree through probing methods. The performance improvement methods, by design, also repaired the tree. Other research focused on addressing the long delays and performance issues by using redundant paths, replicating data or using wholly redundant trees. These methodologies repair the multicast trees as needed in the presence of failures. One form of multicasting uses several multicast trees where data is sent equally on each tree. This methodology is called multiple tree multicasting. Multiple multicast trees have been shown to benefit multicasting applications in that they increase throughput and reliability and several approaches to multiple tree multicasting have been implemented. Multiple multicast trees are built at the application layer to support the data distribution. In a given multicast tree, a subset of the client nodes assists with the data delivery. With multiple multicast trees, more client nodes assist with data delivery. Whether streaming video or sharing files such as with Napster or BitTorrent, using multiple multicast trees is more efficient than using a single multicast tree. These approaches were designed to manage node loss which is a fundamental problem of single multicast trees specifically targeting wireless networks. These mechanisms to manage node loss include additional replication of packets besides just the expected distribution through the tree, forward error correction [14] and multiple descriptions coding [11]. No matter which methodology is examined, repairs of the multicast trees take a long time with respect to the time frame for data delivery where data delivery is on the order of tens of milliseconds and tree repair is on the order of tens of seconds.

Unfortunately, no one multicast solution has addressed all of the problems of multicast data delivery. The ideal solution would provide the following properties so that the application layer would be only minimally affected due to changes to the structure of the multicast tree [1][2][5][7][9][13][18]:

1. *Minimize time to deliver the data.*
2. *Maximize the number of packets delivered.*
3. *Minimize bandwidth utilization.*
4. *Minimize network maintenance overhead.*
5. *Quick detection of failures.*
6. *Quick response to those detected failures.*
7. *Seamless repair mechanisms.*

When there are no disruptions to the multicast overlay network, the data is transmitted effectively and received appropriately by all client nodes via the multicast trees. However, in the presence of network congestion and node turnover, problems arise. Early multicast overlay network research investigated many of the properties of an ideal solution but fundamentally failures still cause too much delay in data delivery. Since people are the ultimate client of the data any perceived quality of experience degradation causes frustration. Although an ideal solution is not possible, any new solution should have most of the following attributes:

1. *It must be efficient in that it performs comparably to other solutions in a static environment and does not cause undue stress on the multicast overlay network by reducing bandwidth utilization and multicast tree maintenance overhead.*
2. *The solution must be resilient in that failures of the multicast overlay network are transparent to the application or minimally disruptive.*
3. *It must be quick to detect network failures and just as quickly respond to the failures to maintain the transparency.*
4. *The solution must provide excellent performance under a variety of conditions that will rival or surpass other similar solutions.*
5. *The solution must provide for quick recovery in the face of network dynamic behavior seamlessly repairing the multicast tree network.*

The main contribution of this work is the development of Probabilistic Multicast Trees as an optimizing mechanism to improve the data delivery latency and data delivery efficiency of *any* multiple multicast tree methodology. PMT was designed to be inserted into any multiple tree multicasting model. The advantage gained by using PMT is that it improves upon the management of the dynamic behavior of the clients where the target connectivity is constantly changing because of its feedback mechanisms and probabilistic tree selection. PMT will be compared against SplitStream, a multiple multicast tree model, under a variety of conditions to show the advantages that PMT provides. SplitStream [6] is a multicasting model that relies on a structured peer-to-peer overlay network called Pastry [10][21], and on Scribe [5], an application-level multicast system built upon this overlay to construct and maintain multicast trees.

3. Probabilistic Multicast Trees

PMT is based on latency feedback. In order to provide latency feedback a separate periodic thread was created that executes at a fixed time period of one second. This thread sends feedback data to its parent for each multicast tree. The feedback packet consists of the averaged feedback from all the parent's children and the parent's average latency delay value. Of course, missing feedback from children causes the averaged delay value to be larger thereby penalizing the multicast tree. New feedback values overwrite older feedback values. It is these feedback values that are used to generate the probability of usage table that the source will use to make a decision about which multicast tree to use for each packet. The Scribe [5] "anycast" functionality was added to enable this feedback from child to parent. The latency feedback mechanism is the key to PMT.

PMT is built upon the following premise: since each multicast tree does not have the same performance characteristics, PMT relies on the latency feedback mechanism from each multicast tree to generate a probability percentage of usage for each multicast tree. The probability percentage of usage for a given multicast tree is a value indicating how frequently a particular multicast tree may be chosen. For each packet sent, one multicast tree is chosen randomly based on its probability percentage of usage. The higher a value for a particular multicast tree, the higher its probability is for being chosen for the next packet to be sent. As a result, the tree with the best performance will be used most often and poorer performance trees will be used less frequently. However, less frequently poorer performance trees will nonetheless occasionally be used possibly yielding improvements in latency feedback possibly due to decreased network congestion for these trees.

There are two reasons for using multiple trees. The first is to maintain the benefits of multiple multicast in that more nodes are actively multicasting the data. The second is to account for changing bandwidth patterns as the underlying networks exhibit their dynamic behavior. The decision to select a multicast tree for a packet about to be sent is based on the generation of a random number and this number is applied against the trees' probability percentage of usage to make the selection. As the performance of the multicast trees change due to node loss, network congestion, tree performance improvement or other changes due to mobile nodes, the latency feedback mechanism continually provides updated latency values to the source so that as the multicast trees' probability percentage of usage is recalculated tree selection chooses the best tree most often at any given time. Recalculation is performed at regular intervals once per second.

PMT improves upon the management of the dynamic behavior of the clients when the target connectivity is constantly changing because of its feedback mechanisms and probabilistic tree selection. This improvement manifests itself in data delivery latency, a metric measured as an output of the process. An improvement in the metric is an indication that using PMT is advantageous.

Figure 1 illustrates three multicast spanning trees. To the source node each tree is a wholly separate multicast tree. In SplitStream each tree is used in a round robin fashion to send each individual packet. For example, the first packet is sent on the blue tree, second packet is sent on the red tree, the third packet is sent on the black tree. The fourth packet will be sent on the blue tree as the process repeats until all the data is transmitted. Figure 2 shows the three non-overlapping trees.

PMT does not follow this round robin process for tree selection. For this example, Tree *2* has been determined to be a more efficient tree for transmission than Tree *1*. Tree *1* has been determined to be a more efficient tree for transmission than Tree *3*. Tree *2* is assigned a probability of usage of 0.67 based on its relative efficiency as compared to the other two trees. Tree *1* is assigned a probability of usage of 0.31 based on the same criteria. Tree *3* is assigned a probability of usage of 0.02. The efficiency of each tree was measured via feedback over a period of time with the network in a steady state mode which resulted in the assigned probabilities.

The calculation of the probabilities will be described below. To choose a tree for transmission a random number is generated. If the random number is less than 0.67 then Tree *2* is chosen. If the random number is between 0.67 and 0.98 then Tree *1* is chosen. If the random number is greater than 0.98 then Tree *3* is chosen. This process is repeated for each packet transmitted. As long as no significant changes occur in the performance of the trees, then the probability of usage for each tree will remain the same. When the efficiency of the trees changes then the probability of usage will change based on the relative performance of each tree.

PMT data delivery latency includes the estimated missing packet latency delay times. Also, congestion has much less impact on the latency feedback when compared to lost node impact. This is to be expected since the missing node penalty value is so large compared to congestion.

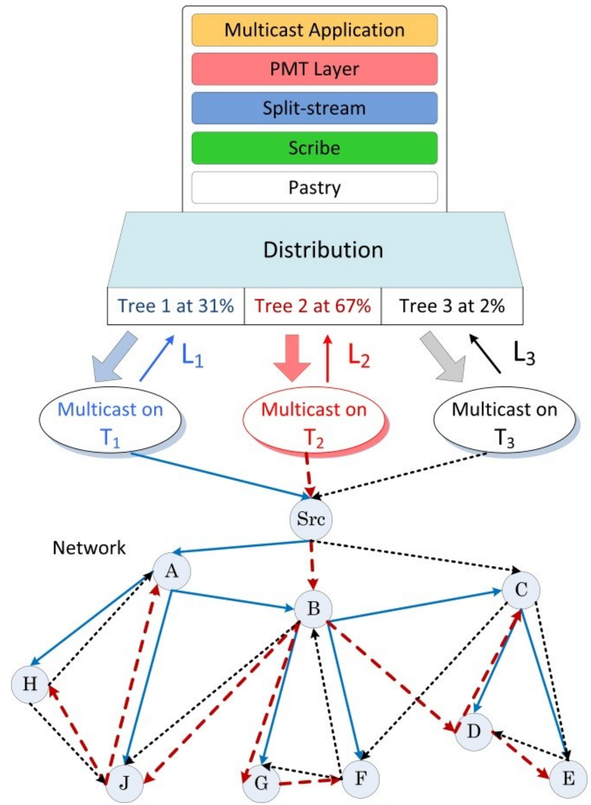


Fig. 1 PMT Multicast Tree Selection

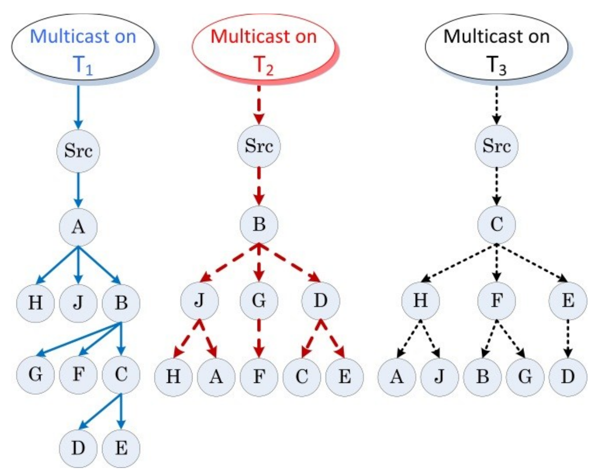


Fig. 2 Three Multicast Trees

4. Data Metrics

Simulation data is collected and passed through a series of calculations that will be used for analysis and comparison.

One metric is *total delay latency time*. This is the summation of all the time differences from all the packets received. During dynamic testing some nodes are lost which means that the remaining nodes will not receive all of the packets. The *non-normalized data delivery latency* is the total delay latency time for the actual number of packets received and does not include the missing packet latency delays.

The *normalized data delivery latency* metric, *NL*, is based on the non-normalized data delivery latency metric. NL is calculated as if the receiving node had received every packet. For each lost node a sufficiently large penalty value is substituted for the feedback value from the node.

A *Better/Worse percentage* metric, *BW*, is calculated by dividing the PMT NL value by the SplitStream NL value.

$$BW = \frac{NL_{PMT}}{NL_{SplitStream}} \times 100\% \qquad (3)$$

A value less than 100% means that the PMT method performed better than SplitStream. The BW percentage is an indication of how well the PMT method performed in a simulation test. The results of these calculations are analyzed statistically and presented in the next section.

5. Results

Simulations were run as follows. The source puts the tree number into the packet and a time stamp into each packet. The receiver uses the tree number to drive metric collections for each tree and it performs a difference calculation to generate the delay time from the source. This delay time is added to the data delivery latency total and is used for worst case delay comparison. The following enumeration describes the raw data collected by the nodes of the multiple multicast networks.

1. *The source node tracks the number of packets sent on each tree. For SplitStream this will always be the same number; however, for PMT, the number will vary for each tree.*
2. *The client node tracks the total number of packets received on each tree.*
3. *The client node tracks the worst case data delivery latency per tree.*
4. *The client node tracks the total data delivery latency for all packets received.*

PMT is compared against SplitStream using the total delay latency metric. Each set of tests is averaged and the mean of the total delay latency is compared directly.

Each dynamic test simulation run had an initial node count and a specified number of nodes to be removed - approximately 10% of the total. After all the nodes were created a subset of nodes were randomly chosen for removal beginning 8 seconds after data transmission started. The chosen nodes were removed from the "active" list and placed on the "to-be- removed" list. The "to-be-removed" node list is iterated to remove the nodes at the appropriate point in the simulation so that the nodes can be removed from the trees. This process provided sufficient dynamic behavior for comparison.

GT-ITM (Georgia Tech Internet Topology Models) is an internet topology generator [20]. Since its release GT-ITM has been widely used in the scientific community for network simulations. We used the GT-ITM model with 8 trees and node counts of 550, 1100, and 2200. Figure 3 and Figure 4 show the average total data delivery latency for the two methods, PMT and SplitStream. As indicated from the means of the two charts, PMT shows a 16% improvement in average total data delivery latency. This improvement percentage is actually more impressive because the SplitStream data delivery latency was calculated using the non-normalized formula which does not include the missing packet latency delays whereas the PMT means included the estimated missing packet latency delay times. The forth Figure 5, shows the calculated comparisons between the actual data delivery latency for PMT and a calculated normalized data delivery latency for SplitStream within the bounds of the same network configuration. The comparison shows a 14% improvement by the PMT over the SplitStream code.

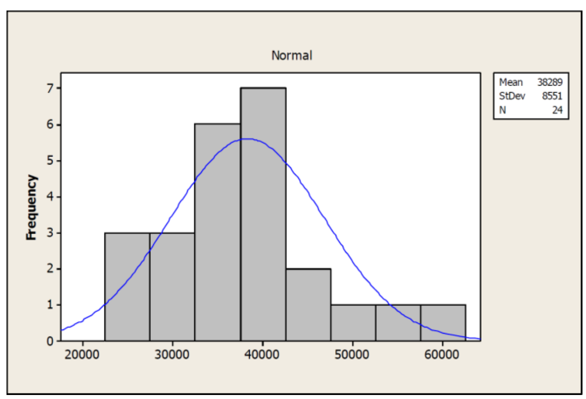


Fig. 3 SplitStream Data Delivery Latency

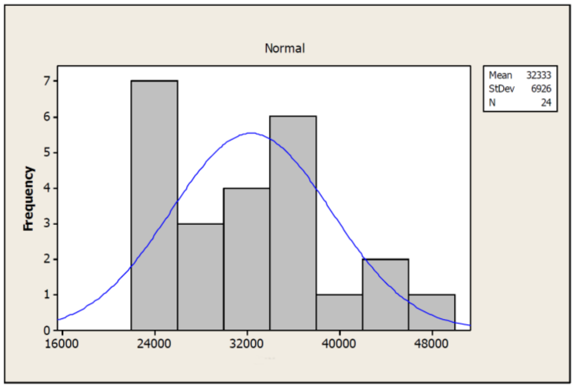


Fig. 4 PMT Data Delivery Latency

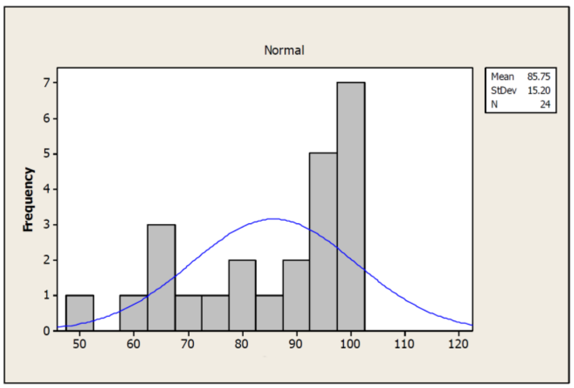


Fig. 5 PMT versus SplitStream Better/Worse

6. Conclusions

This paper has presented PMT, an optimizing mechanism that is intended to improve the capabilities of *any* multiple multicast tree methodology with respect management of node loss and network congestion. Simulations with PMT have shown to improve data delivery latency over the multiple multicast tree scheme SplitStream. As a byproduct data delivery efficiencies are improved by PMTs avoidance of trees with high node loss.

References

[1] D. Andersen, H. Balakrishnan, F. Kaashoek and R. Morris, "Resilient Overlay Networks," Proc. 18th ACM SOSP, October 2001.

[2] T. Baduge, A. Hiromori, H. Yamaguchi and T. Higashino, "A distributed algorithm for constructing minimum delay spanning trees under bandwidth constraints on overlay networks," Systems and Computers in Japan, Vol. 37, No. 14, pp. 15 – 24, 2006.

[3] S. Banerjee, S. Lee, B. Bhattacharjee and A. Srinivasan, "Resilient multicast using overlays," Proc. of ACM SIGMETRICS, June 2003.

[4] S. Birrer and F. E. Bustamante, "Magellan: performance-based, cooperative multicast," 10th International Workshop on Web Content Caching and Distribution (WCW 2005), pp. 133- 143, Sept. 2005.

[5] S. Birrer and F. E. Bustamante, "Resilient peer-to-peer multicast without the cost," Proc. of MMCN, January 2005.

[6] M. Castro, P. Druschel, A.-M. Kermarrec, A. Nandi, A. Rowstron, and A. Singh, "Splitstream: High-bandwidth multicast in cooperative environments," Proc. of the 19th ACM SOSP, October 2003.

[7] M. Castro, P. Druschel, A.-M. Kermarrec and A. Rowstron, "SCRIBE: a large-scale and decentralized application-level multicast infrastructure," IEEE Journal on Selected Areas in Communications, vol. 20, no. 8, pp. 1489–1499, 2002.

[8] Y. Chu, S. G. Rao, S. Seshan and H. Zhang, "A Case for End System Multicast," IEEE Journal on Selected Areas in Communications, vol. 20, no. 8, Oct 2002

[9] N. Feamster, D. Andersen, H. Balakrishnan, and F. Kaashoek, "Measuring the Effects of Internet Path Faults on Reactive Routing," Proceedings of ACM International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS), 2003.

[10] Z. Fei, and M. Yang, "A Proactive Tree Recovery Mechanism for Resilient Overlay Multicast," ACM Transactions on Networking, vol. 15, no. 1, February 2007.

[11] V. K. Goyal, "Multiple Description Coding: Compression meets the Network," Signal Processing Magazine, vol. 18, no. 5, pp. 74-93, 2001.

[12] D. A. Johnston, D. R. McIntyre, F. G. Wolff and C. A. Papachristou, "Optimizing Application Level Multicast Trees over Wireless Networks," IEEE NAECON 2011.

[13] D. A. Johnston, D. R. McIntyre, F. G. Wolff and C. A. Papachristou, "Probabilistic Multicast Trees," Journal of Computer Science & Technology (JCS&T), vol. 12, no. 1, April 2012.

[14] D. Koutsonikolas, Y. C. Hu, "The Case for FEC-based Reliable Multicast in Wireless Mesh Networks," Proc. of 37th Annual International Conference on Dependable Systems and Networks Washington DC, pp 491 – 501, 2007

[15] A. Rowstron and P. Druschel, "Pastry: Scalable, decentralized object location and routing for large-scale peer to peer systems," Proceedings of the 18th IFIP/ACM International Conference on Distributed Systems Platforms (Middleware'01), 2001

[16] J. H. Saltzer, D. P. Reed and D. D. Clark, "End-to-End Arguments in System Design," ACM Transactions on Computer Systems, vol. 2, no. 4, pp 277-288, Nov 1984

[17] K.K. To and Jack Y.B. Lee, “Parallel Overlays for high data-rate multicast data transfer Computer Networks,” 51, pp. 31-42, 2007

[18] D. Tran, K. A. Hua and T. Do, “ZIGZAG: An Efficient Peer to Peer Scheme for Media Streaming,” Twenty-Second Annual Joint Conference of the IEEE Computer and Communications (INFOCOM 2003), Vol. 2, pages 1283- 1292, 2003

[19] V. Venkataraman, K. Yoshida and P. Francis, “Chunkyspread: Heterogeneous Unstructured End System Multicast,” Proceedings of 14th IEEE International Conference on Network Protocols, 2006.

[20] E. W. Zegura, K. L. Calvert, and S. Bhattacharjee, “How to Model an Internetwork”, Proceedings of IEEE INFOCOM ’96, San Franscisco, pp. 594–602, March 1996.

[21] FreePastry Simulation http://www.freepastry.org/FreePastry

David A. Johnston received a Ph.D. in Engineering and Computer Science from Case Western Reserve University.

David R. McIntyre received a Ph.D. in Computer Science from the University of Waterloo. He is Associate Professor of Computer Science at Cleveland State University.

Francis G. Wolff received a Ph.D. from Case Western Reserve University. He is a Senior member of the IEEE and the Senior member of the ACM.

Christos A. Papachristou is Professor of Electrical Engineering and Computer Science at Case Western Reserve. He received the Ph.D. degree in Electrical Engineering and Computer Science from Johns Hopkins University. He is a Fellow of the IEEE and a member of the ACM and Sigma XI, and is listed in Who's Who in America.

Detection of Syntactic Aspect Interaction in UML State Diagrams Using Critical Pair Analysis in Graph Transformation

Zaid Altahat[1], Tzilla Elrad[2] , Luay Tahat[3] and Nada Almasri[3]

[1] GE Healthcare
Chicago, IL. USA

[2] Illinois Institute of Technology (IIT)
Chicago, IL. USA

[3] MIS Department, Gulf University for Science and Technology
Mishref, Kuwait

Abstract

Aspect Oriented Modeling separates crosscutting concerns by defining Aspects and composition mechanisms at the model level. Composition of multiple Aspects will most likely result in more than one Aspect matching the same join points. Consequently, Aspects do not always interact in a predictable manner when woven together. Intended interaction among aspects is designed by the system designer. Unintended interaction (or interference) must be automatically managed. When the woven aspect demonstrates a behavior that is different than its autonomous behavior, then this is a potential interference. Interference has been recently reported in Aspect Oriented Software Development (AOSD) by the industry. Leaving this problem unsolved may result in erratic software behavior and will hinder the adaptation of AOSD by the industry. This identified problem is similar to a phenomenon that exists in graph transformation systems where multiple Graph Transformation rules share some conflicting elements, it is referred to as Critical Pair Analysis and it provides an algebraic-based mechanism to detect and analyze the interaction of the rules. In this paper we propose a framework to detect unintended Aspect interaction at the model level. The proposed framework transforms Aspects modeled in UML State Diagram to Graph Transformation Rules, and then it applies Critical Pair Analysis to detect unintended interactions among aspects. This will enable developers to specify only the order of precedence for intended interaction among aspects without the need to manually investigate unintended interactions for the combinations of every Aspect to every other Aspect in the system. The proposed interaction detection solution is automated, modular, and independent of the base model; which adds the advantage of not having to re-evaluate the interaction each time the base model changes.

Keywords Aspect Oriented Software Development; Aspect Oriented Modeling; Aspect Interaction; Critical Pair Analysis.

1. Introduction

Software modules are added to other software modules and other components in an incremental way to build software products. This process will most probably result in interaction among the software modules. Most software modules have several complex interactions with other software module through their life cycle. Interference among independently built software modules plays a critical role in undermining the stability of the software product under development [10, 16, 20, 40, 41].

The modular approach of software development that is followed by Aspect Oriented Software Development AOSD [42] makes susceptible to the interference problem among the independently built aspects. The behavior of the software may be unpredictable due to the interference problem. In order to have a predictable behavior by the product, the interference must be eliminated. The elimination of the interference requires a practical approach of detecting all possible interfering aspects. AOSD modularizes concerns that crosscut with other concerns into aspects, which are later woven to the rest of the base system's program (or model). Woven aspects do not usually work in isolation; aspects collaborate to deliver a service. Desired cooperation between aspects is manually designed into the system. Undesired, or unplanned, cooperation on the other hand must be precisely defined using an automated approach. The interference (or unintentional interaction) among the software modules may result in a random behavior of the software product that is dependent on the weaving order. We believe that the interference that might take place among aspects must be automatically and efficiently detected.

This interaction was investigated earlier on the telephony systems and referred to as Feature

Interaction [3, 6, 10, 13]. Different mechanisms [6, 10, 13, 34] were proposed to handle the FI problem. Aspect Oriented Software Development (AOSD) [35] builds software systems by composing crosscutting concerns in a similar approach to the features in the telephony systems. This leads to the Aspect Interaction (AI) problem that is very similar to the FI problem. The AI is not necessarily harmful [25]. But the term AI usually refers to the unintended interaction. If the interaction is planned, order precedence needs to be defined [20]. If a dependency between two Aspects is not planned, then unless an AI detection mechanism is used, the dependency might slip undetected with potential harm to the system. The Motorola WEAVR [17] has reported the AI problem in the Telecomm industry, where precedence is defined for interacting Aspects [20]. Graph Transformation (GT) systems have developed a mechanism to detect conflicts among GT rules [38].

In this paper, we propose a graph-based framework to detect unintended interaction among Aspects in UML State Diagrams. Our approach uses Critical Pair Analysis (CPA) which is a technique originally used in telephony systems to detect features interactions. In our approach we transform Aspects defined in UML state diagrams to GT rules, and then we apply CPA to automatically detect interacting aspects.

This paper is organized as follows; Section 2 describes the different types of Aspect interactions. Section 3 presents the GT systems and CPA. Section 4 presents our proposed framework with an example of an ATM modeled as State Diagram. In the case study presented in Section 5, multiple Aspects are defined and their interactions are analyzed and classified using the generated CPA report. Related work is discussed in section 6. Conclusion and future work are discussed in section 7.

2. Aspect Interaction

With the use of AOSD to manage separation of concerns, AI is an inevitable issue. AI takes place when multiple Aspects share conflicting elements in their pointcuts or advices. Multiple aspects are said to be independent if the order of applying aspects result in the same model. Two models are considered syntactically the same if there is a bijective mapping between the two models. That is given two models M1 and M2, for each element in M1 there is one element in M2 with the same properties. There is also the same reverse mapping from M2 to M1.

Interaction among Aspects exists in the form of either dependency or conflict. This kind of AI is referred to as Aspect-to-Aspect interaction [20]. Even non-conflicting and independent Aspects might have unintended impact on the structure of the base model; this kind of interaction is referred to as Aspect-Base interaction. Aspects may also have unintended impact on the behavior of the base model, this kind of interaction is referred to as *semantical* interaction [5]. The next 4 definitions will shed light on the different types of interactions. Let:

$\mathbf{M_1}$ = The result of applying Aspect A1 to the Base Model (BM).
$\mathbf{M_2}$ = The result of applying Aspect A2 to the BM.
$\mathbf{M_{12}}$ = The result of applying A1 *then* A2 to the BM.
$\mathbf{M_{21}}$ = The result of applying A2 *then* A1 to the BM.

Definition 1: Two Aspects do not have interaction between them iff $\mathbf{M_{12}} = \mathbf{M_{21}}$.

Definition 2: A *dependency* exists between two Aspects if ($\mathbf{M_{12}} = \mathbf{M_2}$) or ($\mathbf{M_{21}}=\mathbf{M_1}$)

Definition 3: A *conflict* exists between two aspects if ($\mathbf{M_{12}} = \mathbf{M_1}$) or ($\mathbf{M_{21}} = \mathbf{M_2}$)

Definition 1 states that, regardless of the order of applying Aspects, the output model is the same. This is only possible if the application of one Aspect does not alter the applicability of the other Aspect. If $\mathbf{M_{12}} \neq \mathbf{M_{21}}$, then AI exists between A1 and A2 in the form of either dependency (definition 2) or conflict (definition 3). If ($\mathbf{M_{12}} \neq \mathbf{M_{21}}$ *And* $\mathbf{M_{12}}=\mathbf{M_2}$) then A1 depends on A2, or if ($\mathbf{M_{12}} \neq \mathbf{M_{21}}$ *And* $\mathbf{M_{21}}=\mathbf{M_1}$) then A2 depends on A1. A conflict is defined as either ($\mathbf{M_{12}} \neq \mathbf{M_{21}}$ *And* $\mathbf{M_{12}}=\mathbf{M_1}$) or ($\mathbf{M_{12}} \neq \mathbf{M_{21}}$ *And* $\mathbf{M_{21}}=\mathbf{M_2}$), which means A1 disables A2, or A2 disables A1, respectively.

3. Graph Transformation and Critical Pair Analysis

This section describes Graph Transformation (GT) and Critical Pair Analysis (CPA) which are used in telephony systems to detect features interaction (FT). A graph transformation applies a GT rule (**P** = **L, R**) to a host graph G; where **P** is a production, **L** is the left hand side (LHS) graph, and **R** is the right hand side (RHS) graph. **P** may also have a set of Negative Application Conditions (NAC), which are elements that may not exist for a rule to apply. A GT rule replaces graph **L** with **R** in host graph G.

Figure 1 shows a simple graph, referred to as host graph, (left), a GT rule (middle) with its LHS (**L**) and its RHS (**R**) components, and the generated graph (right). Host graph is searched for a graph morphism of ***L***, referred to as a *match*. A graph morphism between two graphs, G and H, is a bijective mapping (ψ) between the vertices of G and the vertices of H, such that two vertices u and v are adjacent in G iff their mapping vertices ψ(u) and ψ(v) are adjacent in H. If a match is found, the graph ***R*** is applied to the host graph. GT works as follows:

- Elements in *L* and in *R* are preserved in the generated graph.
- Elements in *L* but not in *R* are deleted from the generated graph.
- Elements in *R* but not in *L* are created in the generated graph.

Figure 1-(middle) shows a NAC edge between the states '**b**' and '**c**' in the LHS, marked with '**X**'. The NAC will be used in transforming some of the pointcut constructs, such as 'XOR'. With out the NAC edge in Figure 1, the matching mechanism will only check for the existence of vertices '**b'** and '**c'** without checking the absence, or presence, of an edge between '**b**' and '**c**'. These requirements for morphism come from L. According to **R** edge '**e4**' is created and vertex '**d**' is removed. Generated graph is presented in Figure 1 (right).

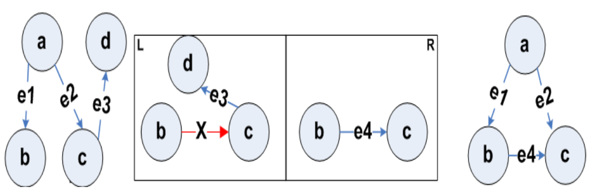


Figure 1. A graph transformation rule on a directed labeled host graph.

GT rules are used to apply changes to a host graph in order to integrate a new feature. Two GT rules that overlap are said to be critical pair. Critical Pair Analysis (CPA) is used to detect conflicts and dependencies in GT Systems. Two rules are in conflict if one rule disables another rule. On the other hand, two rules are dependent if one rule enables the other. Critical Pair Analysis (CPA) [23] is "a pair of transformations both starting at a common graph G such that both transformations are in conflict, and graph G is minimal according to the rules applied." [32]. That is the GT rules **P1** and **P2** form a critical pair if both, **P1** and **P2**, can be applied to the same minimal graph G. But applying **P1** will prohibits the application of **P2** and/or vice versa. Certain tools, such as Attributed Graph Grammar (AGG) [1], provide graph transformations and CPA. An attributed graph allows the definition of attributes on graph elements. CPA and NAC [38] are combined to detect conflicts in GT systems.

4. Aspect Interaction Detection Framework

Aspect Oriented Modeling (AOM) [19] follows an approach similar to the GT systems by querying and adapting base model elements. If a mapping is created between GT rules and Aspects in AOM, then the CPA technique can be used to analyze AI to detect any conflicts and dependencies among Aspects

Since applying Aspects to a base model involves matching and modifying elements in the base model similar to those of the GT systems, our approach uses CPA to detect syntactic interaction among aspects in the UML State Diagrams. UML State Diagrams are increasingly used in modeling wide range of embedded devices, from small gadgets to Telecom Systems [17].

The proposed framework detects potential conflicts and dependencies among Aspects without the need to check the base model for the pointcuts applicability. It achieves this by inspecting all combinations of the pointcuts and advices of all Aspects in a pair-wise manner. This approach will report potential interaction among all Aspects, even if some Aspects might not have a match and regardless of the base model. The advantage of this approach is to avoid regenerating the AI report each time the base model changes.

The framework has a tool which automatically transforms Aspects modeled in UML state diagram to GT rules. Additionally, it uses AGG tool [1] to generate interaction report based on CPA analysis applied on the provided GT rules. **Figure 2 Framework Structure**shows the architecture of the framework where detecting aspect interaction is done in two phases. In Phase one, Aspects are transformed to GT rules automatically using Aspect-to-GT tool which was developed. In phase two, the GT rules obtained from phase 1 are fed to AGG tool which applies CPA analysis and generates the interaction report. The generated report has two matrixes, one showing the minimal conflicts between all pairs of Aspects, and the second matrix shows the minimal dependencies between all pairs of Aspects.

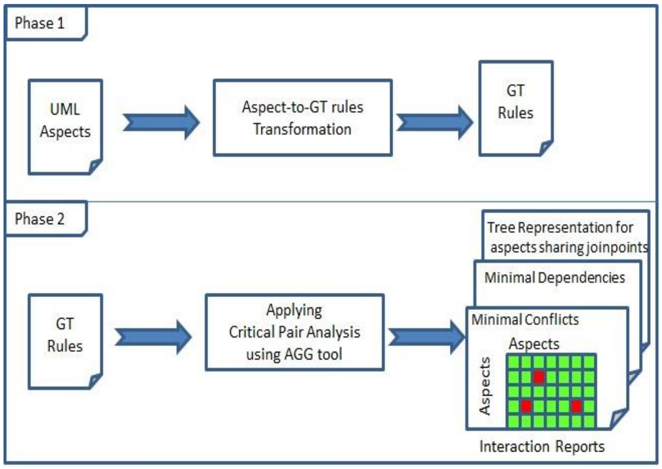


Figure 2 Framework Structure

In section 4.1 we explore the main characteristics of our proposed solutions. In section 4.2, we demonstrate how Aspects modeled in UML state diagram are transformed to GT rules, and in section 4.3 we explain how AGG tool is used to generate CPA interaction reports.

4.1. Characteristics of the Framework

Our proposed solution to the aspect interference problem provides an interference detection solution that is 1) Automatic, does not require the user's intervention, 2) Modular, analysis is done independently of the base system, 3) efficient, the performance of the approach is reasonable, and 4) Practical, does not require extra level of expertise. Aside from the characteristics of our proposed solution, using our solution may eliminate the extremely laborious and near impossible manual process of identifying interfering aspects and increase the confidence in AOSD and bring it one step closer to the industry.

Automation and modularity are essential to any interference detection approach. An approach is modular if the detection analysis is performed on the aspects and the base model separately. Using the proposed approach, users will be able to identify interacting Aspects independently of the base model. This way users need to define only order precedence for identified Aspects once, when they are defined, not each times the model changes. So changes to the base model will not affect the AI, only changing Aspects will require a re-run of the analysis.

Without Automation for Aspect interactions, Aspects designers would have to specify order precedence for all Aspects in the system. This requires large efforts parts of which are useless and wasted. Aspects' designers may still be interested in specifying order for certain Aspects, but they do not have to specify it for all Aspects.

Currently our proposed framework studies the Aspect-to-Aspect interaction; the other types of interactions discussed in section 2 are planned for future work.

4.2. Transformation of Pointcuts to Graph Transformation Rules

In this section we present an example that will demonstrate how Aspects with pointcuts consisting of composite state and compound transitions are transformed to some GT rules. The example consists of an ATM machine described by the UML State Diagram presented in Figure 3. The ATM lacks the behavior to diagnose and early terminate the ATM machine. The behavior is added to the base ATM model by the 'diagnostic' concern, presented in section 4.2, that has 4 Aspects.

Figure 3 presents the UML State Diagram for a bank ATM. Since the *Active* state is composite non-orthogonal, only state '*validating'* will have the incoming transition '*card_in*'. The *Maintenance* state is orthogonal, so both states '*testing'* and '*waiting'* will receive the incoming transitions '*maintain'*.

In order to make it easier on the reader to follow, we numbered each state in Figure 3 and used the numbers in the generated GT rules. Vertices whose names are separated by a '|', for instance the vertex '**1|4**', represent substates in the composite orthogonal state *Maintenance*. Digits to the left of '|' come from the top region, and digits to the right come from the bottom region. When the state '*Maintenance'* becomes active, states '*testing' **(1)*** and '*waiting' **(4)*** become active.

The following 4 Figures, 4 through 7, present the 4 Aspects which are part of the concern 'diagnostic' that will add the behavior to diagnose and early-terminate the ATM. Figure 4-(a) presents the first Aspect A1. Elements marked with '**E**' are exposed and passed to the weaver to adapt. Also to simplify presentation of the GT rules, if an element is presented in the LHS but not in the RHS, it does not mean that the element is deleted, they are just not shown for simplicity.

The Aspect in Figure 4-a is transformed to three different GT rules shown in Figure 4-b. The three vertices (2|6), (2|5), and (2|4) represent the different states the composite state 'Maintenance' might be in while in state 'self_diagnostic' (2). The RHS of the GT rules in Figure 4-b represent the creation of the

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 1, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

edge ‘diagnostic’ between the states ‘validating’ (9) and ‘self_diagnostic’ (2).

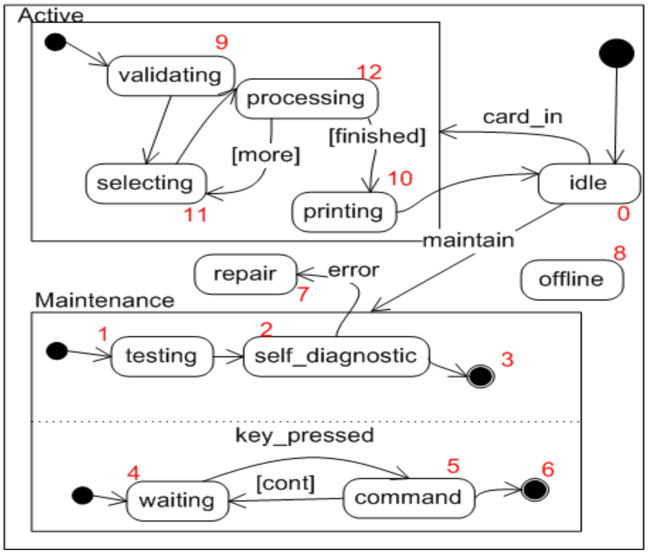


Figure 3 UML State Diagram of an ATM

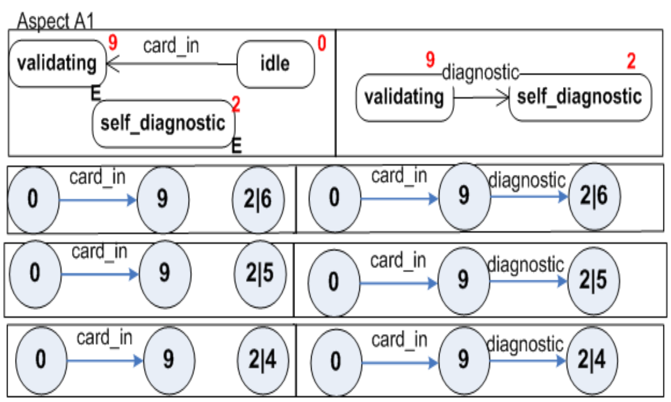


Figure 4 a-(top) Aspect 1, b-(bottom) GT rules

Figure 5 presents the second Aspect A2. The advice of the Aspect creates the edge ‘eject’ to the sequential state ‘Active’. Every substate in the state ‘Active’ will have an incoming edge labeled ‘eject’ from the state ‘idle’. The vertices 9,10,11, and 12 of the generated GT rules in Figure 5-b represent the substates of state ‘Active’. Note, to simplify the presentation of the GT rules, the RHSs do not show the edge ‘diagnostic’ and the vertices (2|6), (2|5), and (2|4) which are preserved in the host graph.

Figure 6 presents the third Aspect A3. The Aspects creates the *fork* transition ‘diagnostic’ which forks to the two sub states, ‘self-diagnostic’ and the final state of the bottom region in the composite state ‘Maintenance’. These results in one GT rule are shown in Figure 6-b.

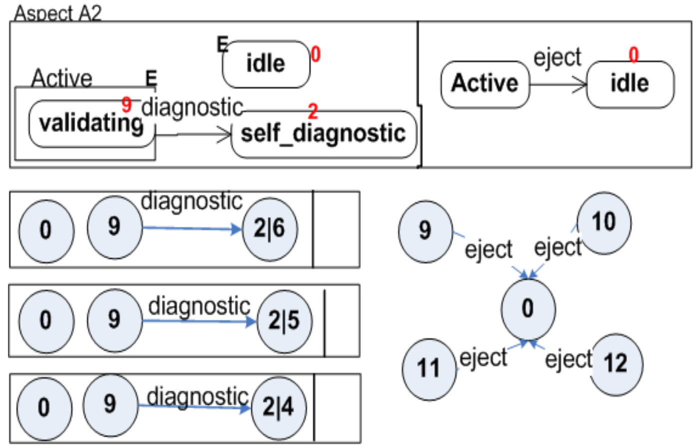


Figure 5 a-(top) Aspect 2, b-(bottom) GT rules

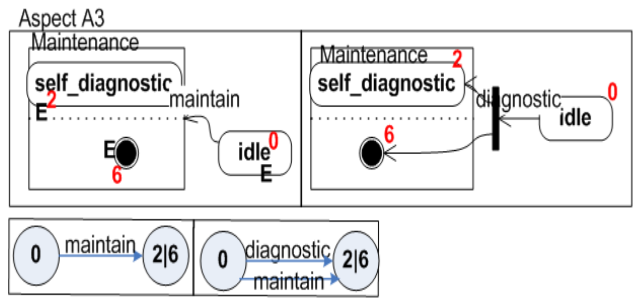


Figure 6 a-(top) Aspect 3, b-(bottom) GT rules

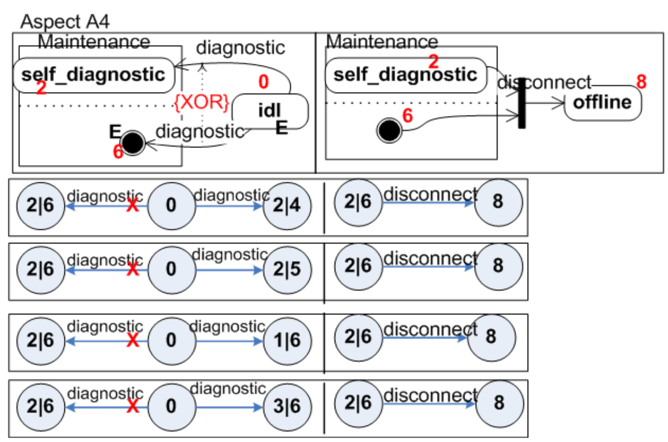


Figure 7 a-(top) Aspect 4, b-(bottom) GT rules

Figure 7 presents the fourth Aspect A4. The pointcut will match the edge ‘diagnostic’ from the state ‘idle’ to the state ‘self_diagnostic’, or from the state ‘idle’ to the state final state of the bottom region, but not from both. The NAC is used to transform the ‘XOR’ element. First two of the 4 GT rules in Figure 7-b present the GT rules that match the transition ‘diagnostic’, and the states ‘idle’ and 'self_diagnostic'. At the same time it does not allow the same transition between the states 'idle' and the *final* substate. The bottom 2 GT rules show the opposite.

4.3. Generating CPA reports

The GT rules of the four Aspects presented in Figures 4-b, 5-b, 6-b, and 7-b are fed to the AGG to generate the CPA report. There are a total of 11 GT rules. To trace the GT rules back to the Aspects, each of the GT rule's name consist of two parts separated by a hyphen. For example the GT rule "A1-R1" represents the first GT rule (R1) of the Aspect A1, presented in Figure 4. The GT set "A1-*" refers to all the GT rules in A1. Any pair of rules in the GT set "A1-*" that is in conflict or dependency with any GT rule in other GT sets, will cause the Aspect A1 to be in the same conflict or dependency as its rule. For instance there is a conflict, Figure 8 (top), between the rules A3-R1 and A4-R1, which causes the Aspects A3, and A4 to be in conflict. This is because the pointcut of A4 doesn't allow for the transitions created by A3 for its pointcut to have a match. Also by inspecting Figure 8 bottom, we can see that A2 depends on A1 for its pointcut to find a match. One thing to mention is that any conflict or dependency within the same GT set is irrelevant and ignored. Note, due to space, Figure 8 shows only part of the report for the interacting Aspects.

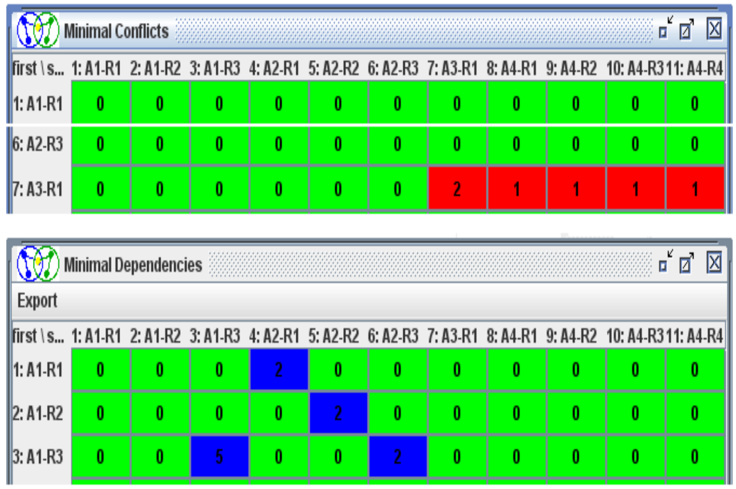
5. Case Study: POTS

In this section, we applied our solution for the detection of aspect interference to the POTS phone system [Kor00] which consists of nearly 40 aspects. Detection of interference is achieved by composing pairs of all aspects, so we have 1600 pairs of aspects. For the 1600 pairs we generate:

- A Tree representation of all the aspects that share a joinpoint.
- A matrix of the conflicts among all aspects.
- A matrix of the dependency among all aspects.

Without our solution to the interference problem, the aspect designer has to manually inspect the 1600 pairs and decide for possible interference among the aspects.

The POTS system consists of the base model and a set of features. The base model is modeled in UML state diagrams. The features were also transformed into aspects and modeled in UML state diagrams. Two types of interference were detected, Dependency among aspects, and Conflicts among aspects.


Minimal Conflicts

first \ s...	1: A1-R1	2: A1-R2	3: A1-R3	4: A2-R1	5: A2-R2	6: A2-R3	7: A3-R1	8: A4-R1	9: A4-R2	10: A4-R3	11: A4-R4
1: A1-R1	0	0	0	0	0	0	0	0	0	0	0
6: A2-R3	0	0	0	0	0	0	0	0	0	0	0
7: A3-R1	0	0	0	0	0	0	2	1	1	1	1

Minimal Dependencies

Export

first \ s...	1: A1-R1	2: A1-R2	3: A1-R3	4: A2-R1	5: A2-R2	6: A2-R3	7: A3-R1	8: A4-R1	9: A4-R2	10: A4-R3	11: A4-R4
1: A1-R1	0	0	0	2	0	0	0	0	0	0	0
2: A1-R2	0	0	0	0	2	0	0	0	0	0	0
3: A1-R3	0	0	5	0	0	2	0	0	0	0	0

Figure 8 The CPA report of the ATM state machine

5.1. The Basic Model

Figures 9 and 10 (in Appendix) present two representations of the same basic call model, UML and graph, respectively. The name of all the states visited during the process of making a phone call start with "o ", for originating, to distinguish them from the receiving phone call states that start with "t ", for terminating.

5.2. The Concerns and their Aspects

The phone system has 10 new concerns defined in 40 aspects. See Altahat [43] for a list of all the aspects. The 10 concerns are typical phone system added features such as call forwarding, call waiting, three way calling, reverse billing, split billing, and few others.

5.3. Results for Aspects with Dependency

Figure 11 (in Appendix) presents the dependency results for the POTS phone system with 40 aspects. The green pairs do not have dependency among them. The shade (blue) pairs on the other hand do have dependency among them and require the designer intervention to at least define the weaving order for these pairs.

We will pick one pair with dependency between them. The aspect CW-S-A7 (Call Waiting - Subscribers), presented in Figure 12 (see Appendix) depends on the aspect CW-S-A5, presented in Figure 13 (see Appendix). The dependency is due to the state CW3:cw hold. The pointcut of the aspect CW-S-

A7 has the state CW3:cw hold. But the state CW3:cw hold does not exist in the base model and only exists after running the aspect CW-S-A5, which creates the state CW3:cw hold.

When inspecting the dependency results in Figure 11 (Appendix), we notice that certain columns (or aspects) are almost entirely shaded (blue). Such as the aspect 1:CFB-S shown in Figure 14 and the aspect 8:RB-S-A1 shown in Figure 15 (Appendix). This means that the aspects has dependency with almost every other aspect in the system.

By inspecting the aspect (1:CFB-S) we notice that the aspect's pointcut matches all states except for the state "idle". Which means that any aspect that creates new states, except for the "idle" state, will have a dependency with the aspect 1:CFB-S. Similarly, the pointcut of the aspect 8:RB-S-A1 matches that states (o o_ hook, o dial tone, o dialed, o busy, o wait for answer, o connected, o wait for onhook, t connected, and t wait for onhook). So any aspect that might create any of these states will have a dependency with the aspect 8:RB-S-A1.

5.4. The Results for Aspects with Conflicts

A conflict between two aspects exists if the application of one aspect prevents the application of the other aspect. Figure 16 (Appendix) presents the results for the POTS phone system with 40 aspects. The green pairs do not have conflict among them. Meaning running them in any order will generate the same model and consequently does not require the designer intervention. The shaded (red) pairs on the other hand do have conflict among them and require the designer intervention to at least define the weaving order for these pairs.

6. Related Work

Several researches has been conducted by the Aspect-Oriented Software Development (AOSD) community in order to reduce complexity and increase reuse of the software by providing modularization of concerns that tend to crosscut. In particular, several researches have been conducted to deal with the problem of Aspect Interaction at different phases of the software development mainly at the requirements level [26, 27, 28, 29]. Khan et al [27] proposed multidimensional concern slicing approach that simplifies identifying requirements dependencies and impact of requirements changes. Dependency graphs for each concern slice are constructed from semi-formal dependency equations, which assist in analyzing the consequential change impact on the requirements. The early determination of requirements dependencies may reduce the undesirable ripple effect propagation. Magno et all [28] studied interactions between concerns and proposed a classification that takes into account the type of concern being studied. Decisional concerns are ranked using a systematic process. Based on this, the orders of concerns' composition were derived. Their approach provides a better understanding of the interactions between all the elements of a system.

S. Ciraci et al [44] proposed graph-base model checking of AOM for Aspect interference. Their approach relies on transforming both the base model and Aspects to Design Configuration Modeling Language (DCML) which is a language proposed by the authors. They run execution simulations to check for Aspect interference. Their work mainly focuses on UML class diagrams and UML sequence diagrams.

Bar-On et al [29] proposed a method that supports the identification of functional requirements that crosscut other functional requirements to generate the derived or modified requirements. To identify crosscutting requirements, they manually used match actions used by requirements and the system modes and states related to these requirements. Their method is based on the observation that, when the same action is used by two requirements it indicates that one of the requirements may crosscut the other.

Shaker and Peters [31] proposed a process for detecting undesirable concern interactions in AO systems at the design phase of the software development process. They describe a statechart weaving language for specifying and later verifying the weaving into an UML model at design stage. Kienzle et al [8] studied Aspect interaction for Aspect-Oriented Programming environment such as AspectJ. They have defined an aspect based on the services it requires from other aspects and on the services it removes. They also established a set of composition rules to solve inter-aspect dependencies. The Aspect Interaction Aspect-Oriented Programming is also classified into different types in [36, 37].

When AOSD was first introduced by Douence et al [14], they studied the interaction problem and proposed a framework for detecting aspect interactions at the language level for AOP. Douence et al considered among the first to look at this problem. Order Precedence for the Aspect-to-Aspect interference of models in the Motorola WEAVR [17] was proposed by Zhang et al [20]. They define three precedence relations as *follows*, *hidden_by*, and *depdendent_on*. Their intent was not to detect interaction, but rather to define precedence relations for interacting Aspects. Mostefaoui et al [5] proposed semantic conflicts between aspects and base model. They translated models to Alloy [2] to be formally verified. Their approach is for semantic verification of aspects and base model interaction. For each aspect they define constrains, pre and post conditions, that will be verified using Alloy at the weaving time.

Bakre and Elrad [33] used Live Sequence Charts to detect AI at the Joinpoint in the form of use-case scenarios. They proposed the Aspect Interaction Charts that build on top of the Live Sequence Charts in order to capture the interactions among various aspects at joinpoints. The Aspect Interaction Charts has the ability to capture aspect interactions at a joinpoint in a common specification in the form of use-case scenarios, and the ability to execute these scenarios while non-invasively manipulating the interactions among the various aspects. They used the tools that come with the Live Sequence Charts language, i.e. the Play Engine to model, view and manipulate aspect interactions at joinpoints.

Havinga et al proposed a graph-based approach [18] to detect composition conflicts due to weaving multiple aspects in AspectJ [22]. They model the structure of programs as graphs and the model introductions as graph transformation rules. They defined explicit rules to describe when composition conflicts related to introductions occur. A prototype tool has been built to detect and visualizes the occurrence of such conflicts in AspectJ programs. The graph-based models are generated automatically from the source code of Java programs and AspectJ introductions. Their approach did not make strong assumptions about either the aspect or base language; it has been designed to be applicable to other AOP languages.

GROOVE [21] is a project centered around the use of simple graphs for modeling the design-time, compile-time, and run-time structure of object-oriented systems, and graph transformations as a basis for model transformation and operational semantics. In their approach they detect pre-defined language violations, such as multiple conflicting method definitions, and cyclic inheritance. The essence of their work is to verify predefined rules in AspectJ, contrary to our approach, which is to detect conflicts among aspects.

Nagy et al [7] proposed a method of analysis of aspect interaction in AOP that was applied to AspectJ. They provide a solution that is constraint-based and declarative for interacting aspects. Their work does not discuss mechanisms for detecting interaction among aspects. Our work concentrates on detecting interaction, dependency and conflicts, of aspects. A mechanism for semantic aspect interaction in Composition Filters for AOP is proposed by Durr et al [11]. They provided a mechanism similar to the mechanisms for detecting deadlock in a computer system. Based on the semantics of the added advices, their approach tries to order aspects in a harmless way.

Mehner et al [9] proposed an approach for analyzing interactions and potential inconsistencies at the level of requirements modeling using variant of UML. Critical Pair Analysis is used to analyze aspect interaction in UML class diagrams. Model transformations are expressed as pre and post conditions that are used in defining graph transformations rules. Pre and post conditions are derived from activity diagrams. In their approach classes and associations among classes are tracked using Attributed Graph Grammar to analyze their interaction. The analysis is performed with the graph transformation tool Attributed Graph Grammar. The automatically analyzed conflicts and dependencies also serve as an additional view that helps in better understanding the potential behavior of the composed system.

Jayaraman et al [12] used Critical Pair Analysis to detect feature interaction in Software Production Lines. Their work presented a graph-based Modeling Aspects using a Transformation Approach to specify how features, modeled in UML, relate to each other. Our framework is also graph based but for UML State Diagrams, in particular composite states and compound transitions and their transformation to GT rules. However, our framework uses CPA technique to detect Aspect-to-Aspect Interaction.

7. Conclusion and Future work

In this paper, we demonstrated how to detect AI in UML State Diagrams. The proposed framework uses Critical Pair Analysis in the GT Systems to detect the interaction, CPA is provided by AGG. The framework has a complexity of $O(n^2)$, where "n" is number of Aspects; but the AI detection for a pair of Aspects needs to be done only once in the system's lifetime. Hence the introduction of a new Aspect to the system will result in (n) pairs between the new and existing Aspects; AI among existing Aspects doesn't need to be reevaluated. Consequently, only a

O(n) is needed for the introduction of a new Aspect. The proposed approach is modular (independent of the base model). This adds a huge advantage in large industrial system.

In order to use CPA, Aspects are transformed to GT rules. KerMeta was used to execute all the model transformations. As seen in section 4, users do not have to define order precedence for all possible combinations of Aspects. Instead user is required only to define order between the Aspect A1 and A2 and precedence between A3 and A4.

However, the proposed framework does not support pattern matching in defining pointcuts, similar to those supported by AspectJ. This is due to the limitation enforced by AGG. There are also other mechanisms that are more expressive, such as Join Point Designation Diagram (JPDD) [4, 15] and the State Machine Joinpoint Model [16] used in the WEAVR [17]. Such mechanisms will result in different GT rules when integrated into our framework. In future work we plan on adding support for JPDD in our framework. As seen in section 4, traceability between the GT rules and Aspects was done manually by using the Aspect#-Rule# naming convention. In large-scale production an automatic traceability is needed which will automatically decide which Aspects are in conflict or dependency without having to report the triggering GT rules.

8. References

[1] http://user.cs.tu-berlin.de/~gragra/agg/

[2] Alloy Homepage. http://alloy.mit.edu.

[3] D. Cansell, D. Méry. Abstraction And Refinement Of Features. Language Constructs for Designing Features. Springer Verlag, 2000.

[4] D. Stein, S. Hanenberg, and R. Unland. “Expressing Different Conceptual Models Join Point Selections in Aspect-Oriented Design. Proceedings of the 5th int’l conference on AOSD. Bonn, Germany. ACM. 2006.

[5] F. Mostefaoui, J. Vachon. “Design-Level Detection of Interactions in Aspect-UML Models Using Alloy”. JOT. Aspect-Oriented Modeling Vol.6, No. 7. August 2007.

[6] TF. BeltaguiT. “Features and Aspects: Exploring feature-oriented and aspect-oriented programming interactions”. Lancaster University. 2003.

[7] I. Nagy, L. Bergmans and M. Aksit, “Composing Aspects at Shared Joinpoints”, In Proceedings of International Conference NetObjectDays (NODe), Erfurt, Germany, pp. 19-38. September 2005.

[8] J. Kienzle, Y. Yu, J. Xiong, “On Composition and Reuse of Aspects”. In Proceedings of the 2nd Workshop on Foundations of Aspect-Oriented Languages (FOAL), Boston, MA, March 2003.

[9] K. Mehner, M. Monga, G. Taentzer. “Interaction Analysis in Aspect-Oriented Models”. 14th IEEE Int’l Req’ Engineering Conference (RE'06) pp. 69-78. IEEE. 2006.

[10] P. Zave. “Feature Interactions and Formal Specifications in Telecommunications”. Vol 26, Issue 8. IEEE Comp Society Press Los Alamitos, CA, USA. 1993.

[11] P. Durr, T. Staijen, L. Bergmans, and M. Aksit, “Reasoning About Semantic Conflicts Between Aspects.” 2nd European Interactive Workshop on Aspects in Software (EIWAS), Brussels, Belgium, September 2005.

[12] P. Jayaraman, J. Whittle, A. M. Elkhodary, and H. Gomaa. Model Composition in Product Lines and Feature Interaction Detection Using Critical Pair Analysis. Springer Berlin, Volume 4735/2007. 2007.

[13] R. J. Hall. “Feature Combination and Interaction Detection via Foreground/Background Models”. TComputer Networks: The Int’l Journal of Computer and Telecom Networking. Volume 32, Issue 4. Elsevier North-Holland, Inc. New York, NY, USA. April 2000.

[14] R. Douence, P. Fradet, M. SüdholtP. P”A Framework for the Detection and Resolution of Aspect Interactions.” Vol 2487/2002. Springer Berlin/ Hedelberg. 2002.

[15] S. Hanenberg, D. Stein, R. Unland. “From Aspect-Oriented Design to Aspect-Oriented Programs: tool-supported translation of JPDDs into Code.” Sixth Int’l Conference on AOSD. Vancouver, British Columbia. 2007.

[16] T. Cottenier, A. van den Berg, T. Elrad. Joinpoint Inference from Behavioral Specification to Implementation. ECOOP. 2007.

[17] T. Cottenier, A. van den Berg, T. Elrad. “The Motorola WEAVR: Model Weaving in a Large Industrial Context.” Sixth Int’l Conference on Aspect-Oriented Software Development. Vancouver, British Columbia. 2007.

[18] W. Havinga, I. Nagy, L. Bergmans, M. Aksit. "A graph-based approach to modeling and detecting composition conflicts related to introductions." AOSD; Vol. 208. Proceedings of the 6th in'l conference on AOSD. Vancouver, British Columbia, Canada 2007.

[19] www.aspect-modeling.org/

[20] J. Zhang. T. Cottenier, , A. van den Berg, J. Gray. **"**Aspect Composition in the Motorola Aspect-Oriented Modeling Weaver". JOT. Special Issue: Aspect-Oriented Modeling Vol.6, No. 7. August 2007.

[21] http://groove.sourceforge.net/groove-index.html

[22] http://www.eclipse.org/aspectj/

[23] H. Ehrig, K. Ehrig, U. Prange, and G. Taentzer. "Fundamentals of Algebraic Graph Transformation." EATCS Monographs in TCS. Springer, 2005.

[24] KerMeta Homepage: http://www.kermeta.org/

[25] F. Sanen, E. Truyen, W. Joosen, A. Jackson, A. Nedos, S. Clarke, N. Loughran, A. Rashid. "Classifying and Documenting Aspect Interactions". Workshop on Early Aspects at AOSD 2006

[26] R. Chitchyan, J. Fabry, L. Bergmans, "Aspects, Dependencies, and Interactions", in Aspect, Dependencies, and Interactions Workshop (held at ECOOP). 2006.

[27] S. Khan and A. Rashid, "Analyzing Requirements Dependencies and Change Impact Using Concern Slicing", in Aspects, Dependencies, and Interactions Workshop,Lancaster, 2006.

[28] J. Magno and A. Moreira, "Concern Interactions and Tradeoffs: Preparing Requirements to Architecture", in Aspects, Dependencies, and Interactions Workshop. Lancaster, 2006.

[29] D. Bar-On, and S. Tyszberowicz, "Derived Requirements Generation" in Aspects, Dependencies, and Interactions Workshop, Germany, 2007.

[30] P. Shaker and D. K. Peters, "Design-level Detection of Interactions in Aspect-Oriented Systems", in Aspects, Dependencies, and Interactions Workshop (held at ECOOP), Lancaster University Computing Department Technical Report Series, ISSN 1477447X. Lancaster, 2006.

[31] F. Munoz, O. Barais, and B. Baudry, "Vigilant Usage Of Aspects" in Aspects, Dependencies, and Interactions Workshop (held at ECOOP), Germany, 2007.

[32] G. Taentzer, "AGG: A Graph Transformation Environment for Modeling and Validation of Software", AGTIVE, US 2003.

[33] S. Bakre, T. Elrad. "Scenario Based Resolution of Aspect Interactions With Aspect Interaction Charts" in the 10th Int'l workshop on AOM (held at AOSD), Vancouver, Canada, 2007.

[34] L. Tahat, B. Vaysburg, and B. Korel. "Requirement Based Automating Black-Box Test Generation." Proc's of the 25th Annual IEEE Int'l Computer Software and Applications Conference, 2001.

[35] http://aosd.net/

[36] M. Rinard, A. Salcianu, and S. Bugrara, "A Classification System and Analysis for AOP". In Proc. 12th Int. FSE-12, Newport Beach, USA, Nov. 2004.

[37] F. Tessier, M. Badri, and L. Badri, "A Model-Based Detection of Conflicts Between Crosscutting Concerns: Towards a Formal Approach". Int. WAOSD, China, 2004.

[38] L. Lambers, H. Ehrig, F. Orejas. "Conflict Detection for Graph Transformation with Negative Application Conditions". Lecture Notes In Computer Science. Germany, 2006.

[39] T. Cottenier, A. van den Berg., T. Elrad. Model Weaving: Bridging the Divide between Translationists and Elaborationists. Workshop on Aspect-Oriented Modeling at the 9th International Conference on Model Driven Engineering Languages and Systems, Milan, Italy, 2006.

[40] Z. Altahat, T. Elrad, L. Tahat. "Applying Critical Pair Analysis in Graph Transformation Systems to Detect Syntactic Aspect Interaction in UML State Diagrams", SEKE08, San Francisco, USA, 2008.

[41] M. Kolberg, E. H. Magill, D. Marples, and S. Rei_. Second feature interaction contest. In "Feature Interaction in Telecommunication ans Software System VI", pages 293-310, May 2000.

[42] www.eclipse.org/aspectj

[43] Z. Altahat, "Detection of Interference Among Aspects in Aspect Oriented Modeling". PhD Thesis, Illinois Institute of technology, 2010.

[44] S. Ciraci, W. Havinga, M. Aksit, C. Bockisch and P. Broek, "A graph-based aspect interference detection approach for UML-based aspect-oriented models." *Transactions on aspect-oriented software development VII*. Springer Berlin Heidelberg, 2010. 321-374.

Appendix : POTS Case Study Figures

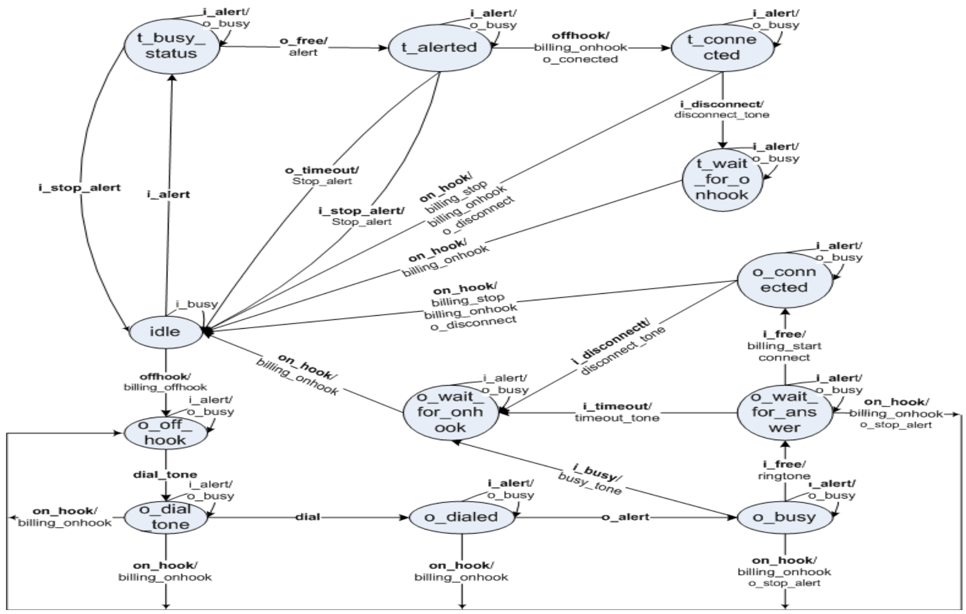


Figure 9The Basic Call Model in UML

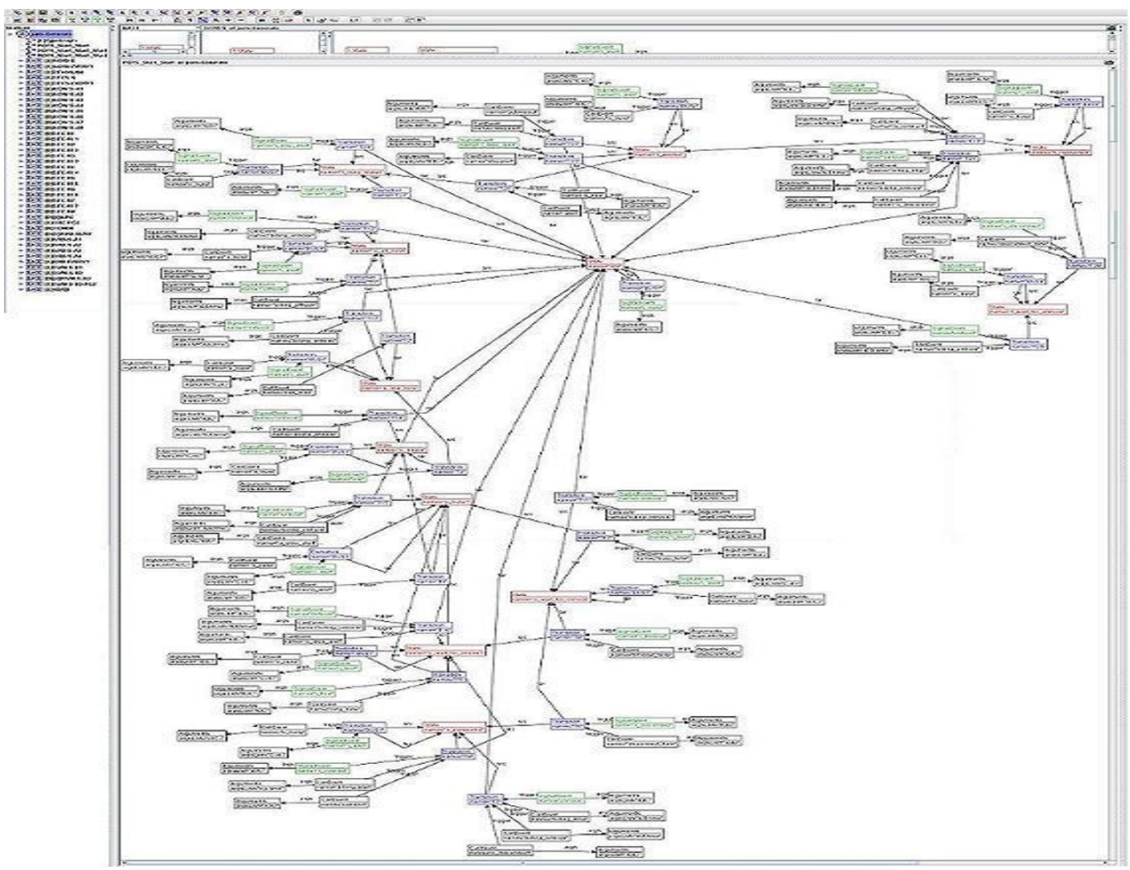


Figure 10 The Basic Call Model in Graph

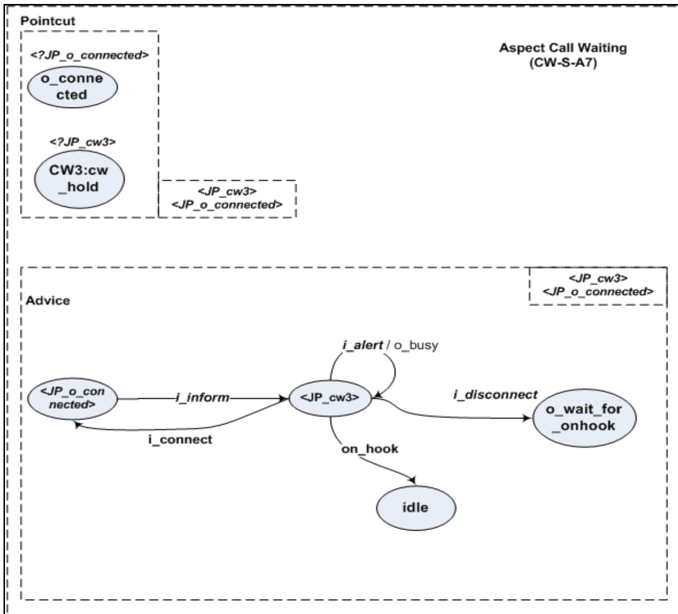


Figure 11 The Aspect Call Waiting for Subscriber (A7)

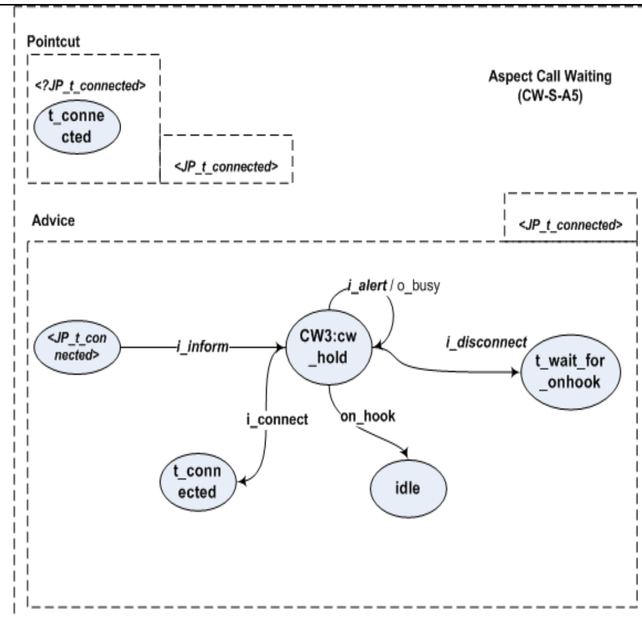


Figure 12 The Aspect Call Waiting for Subscriber (A5)

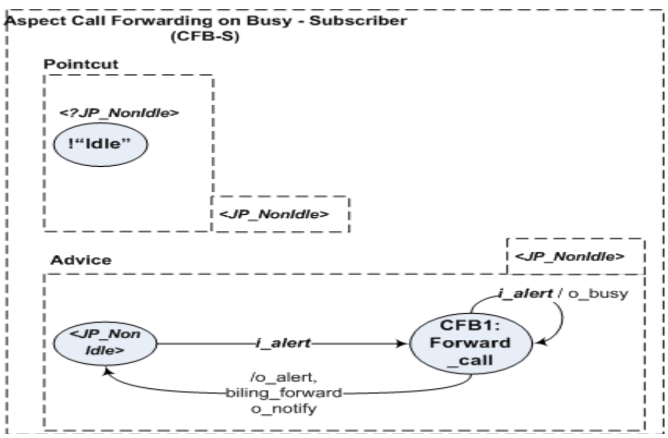


Figure 13 The Aspect Call Frowarding on Busy for Subscriber

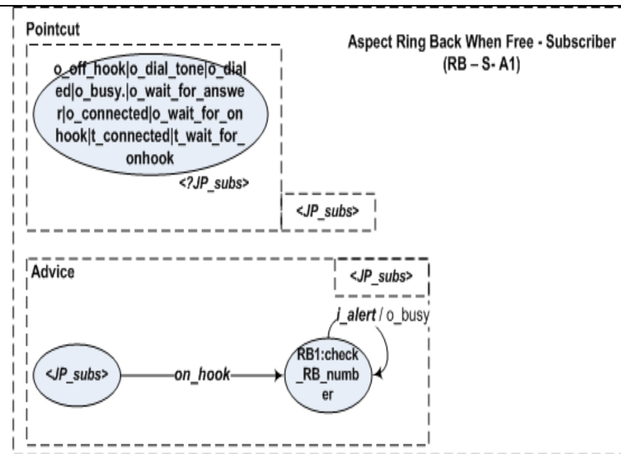


Figure 14 The Aspect Ring Back When Free: Subscribers (A1)

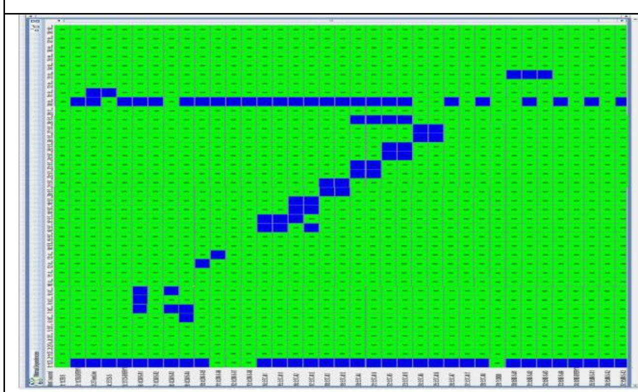


Figure 15 The Results of Dependency Among Aspects

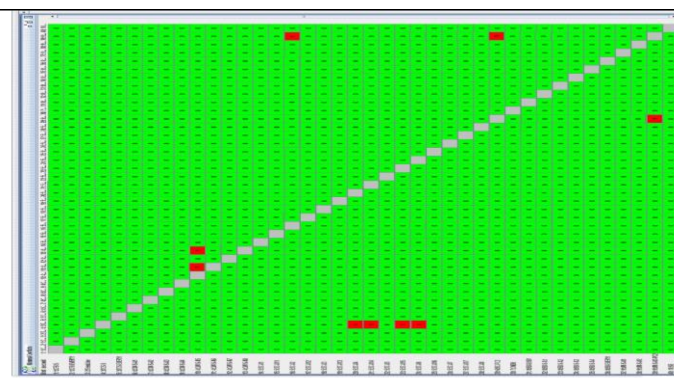


Figure 16 The Result of Conflicts Among Aspects

Simulation Techniques and Control Schemes in Electromyography for A Lower Limb and Power Assist Exoskeleton

Ngoc Khoat Nguyen[1] and Thi Mai Phuong Dao[2]

[1] School of Automation, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China

[2] College of Electrical and Information Engineering, Hunan University, Changsha, Hunan 410100, China

Abstract

This paper represents the first part of our work in which a successful exoskeleton system has been designed. With this design, the concentration is on effects of compliant human – exoskeleton robot on kinematics and muscle activity during human walking with or without carrying loads. In order to achieve these aims, two control methods are used: kinematic and surface electromyography (sEMG) – based control. In this article, three simulation techniques of sEMG signal are investigated. The obtained results of this work are very useful in designing and testing our actual exoskeleton system.

Keywords: *Exoskeleton, Robot, sEMG, Simulation, Control*

1. Introduction

In recent decades, exoskeleton robots have been considered and developed purposely to assist the mobility of physically weak persons who are elder, injured and disabled, or extending the strength of humans, especially in the army [4, 6]. A Lower Limb and Power Assist (LLPA) exoskeleton which has been developing in our laboratory is designed mainly for soldiers who always work under hard conditions. A typical LLPA exoskeleton normally consists of a waist holder, a thigh holder, a lower leg holder, two DC motors, two links, a footrest and two force sensors in one leg [6].

In order to control a LLPA exoskeleton system, many control techniques had been taken up [3, 4] in which the sEMG-based control method has attracted a lot of considerations. According to this method, there are plenty of control trends done to reach remarkable achievements. Nevertheless, few of real time sEMG-based exoskeleton systems are available out of laboratories.

The scope of this present research is to explore deeply the simulation techniques of sEMG signals as the first step of our work to evaluate experimental results of real system. Because researches on sEMG signal- based exoskeleton systems only began a decade ago, models of this signal are still useful for present studies and the future [5, 14]. Most of these models only concentrated on one technique and they were not comprehensive for sEMG signals [3, 13, 15]. This investigation will give three basic models of sEMG signal: physiologically mathematic model, random variable model, and 3-layered volume model. The simulated results are shown clearly to present the survey of sEMG signal which are powerful for LLPA exoskeleton systems.

The second phase of this paper is to analyze two models of sEMG based control scheme of LLPA exoskeleton system: basic and advanced model. These models are evaluated clearly to choose the most suitable control scheme for our system to obtain important design goals of our exoskeleton project as high efficiency and extended application. With the chosen model, we combine a sEMG-based control method with an effective human-exoskeleton robot interaction by using human sensors to measure ankle angles, velocity and acceleration during human walking with or without carrying loads. To implement experiments, we will use an open-circuit respirometer to evaluate human metabolic cost as one of the most important goals of our work.

2. Three sEMG Basic Models

2.1 The Physiologically Mathematic Model

The motor unit in Fig.1 is the smallest activity unit of muscle which is composed by many muscle fibers [8]. The excitability of these muscle fibers through neural control represents a major factor in muscle physiology. To simulate the single fiber potential, we use the volume conduction theory based on the intracellular potential of the fiber [5, 14] which is calculated by (1)

$$e(z)[mV] = 768z^3 \exp(-2z) - 90 \qquad (1)$$

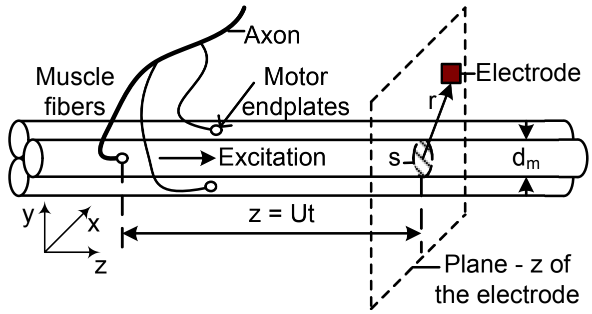


Fig. 1 Motor unit model

where z, in *mm*, is the axial direction. The transmembrane current is

$$i(z)[mA]=\frac{\sigma_i \pi d_m^2}{4}\frac{d^2e(z)}{dz^2} \qquad (2)$$

where d_m, in μm, is the diameter of the fiber, σ_i is the intracellular conductivity ($\sigma_i = 1.01 Sm^{-1}$ [14]).

A relationship between the axial direction and time domain is expressed in (3)

$$z(U,t)[mm]=Ut \qquad (3)$$

where, U, in ms^{-1}, is the propagation velocity of the action potential in the muscle fiber which can be calculated as a function of fiber diameter [5].

$$U(d)[ms^{-1}]=2.2+0.05(d_m-25) \qquad (4)$$

Hence, the single muscle fiber action potential can be yielded by (5)

$$V_{SFi}[mV]=conv[i(t),w(t)] \qquad (5)$$

here *conv* is symbolic of the convolution and *w(t)* is the weighting function which can be calculated by [5]. W. Wang in [15] gave another formula to compute V_{SFi} in (6).

$$V_{SFi}[mV]=-\frac{\sigma_i}{4\pi\sigma_m}\int_S dS\int_{z\min}^{z\max}\frac{\partial e(z)}{\partial z}\frac{\partial(1/r)}{\partial z}dz \qquad (6)$$

where σ_m is the muscle conductivity and r is the distance between the fiber section S and the observation point [5].

Motor unit action potential is obtained in (7) by calculating the total of all the single fiber action potentials

$$V_{MU}[mV]=\sum_{i=1}^{N}V_{SFi} \qquad (7)$$

with N being the number of muscle fibers which can be defined by [5, 14].

2.2 The Random Variable Model of sEMG

The sEMG signals contain important diagnostic data in both the time and frequency domains. In the time domain, root-mean-square (RMS) value and mean rectified value are two common values in which proportional control of myoelectric prostheses is typically dependent on the first value of sEMG. In the second domain, the power spectrum density (PSD) is always used as a common value from which measures of median and mean frequency can be extracted to be used for different control goals.

Sometimes, the sEMG signals can be analyzed as a random distribution which is similar to Gaussian signal [13]. Here, the generation of each motor unit action potential is known as a stochastic process of the status of neuromuscular system and the different impulse delay between two consecutive action potentials. This delay can be used as an independent random variable to model each motor unit action potential approximated to an identical distribution. By applying the central limit theorem, the identically distributed and independent motor unit action potential is reasonable to make sEMG as a pseudo-Gaussian distribution with the roughly equal mean of zero as shown in Fig. 2.

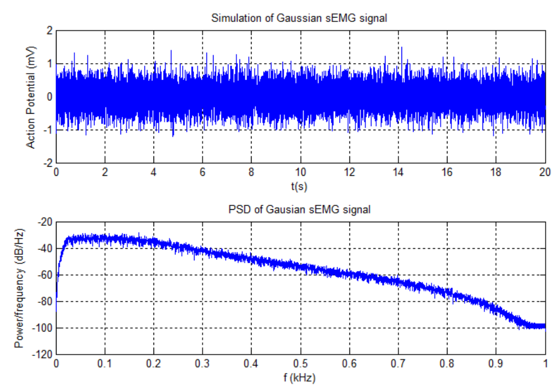


Fig. 2 Simulation of sEMG as a random variable of Gaussian distribution

In order to simulate sEMG signal based on a pseudo-Gaussian distribution, we will use the following formula to calculate the strength of each motor unit action potential:

$$V_{MUi}[mV]=\frac{1}{\sqrt{r^2+(z_{\max}/2-i)^2}} \qquad (8)$$

where r, in *mm*, is the distance between the recording site and muscle fiber, z_{max}, in *mm*, is the length of muscle fiber. Here, we assume that r is a uniformly random variable with a suitable interval from 0.5 to 2 [13].

We also create the sEMG signal by combining a lot of the above motor unit action potentials. To obtain an overall sEMG signal with the representation of time and frequency domains, a bandpass filter with a bandpass filtering of an interval from 20Hz to 200Hz is used in this work.

2.3 A 3-Layered Volume Model for sEMG Signal

Dario Farina and Roberto Merletti in [11] proposed a non-homogeneous (layered) and anisotropic volume conductor model. By combining with Fansan Zhu in [9] , a circuit to simulate sEMG signal based on a 3-layered volume conduction model is presented as shown in Fig. 3.

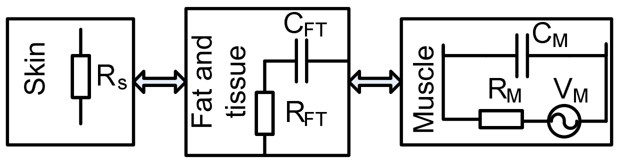


Fig. 3 A 3-layered volume model

This model includes three layers:

(a) the skin layer: is shown by simple resistors R_S;
(b) the fat and tissue layer: is approximated as a low pass filter ($R_{FT}C_{FT}$ filter);
(c) the muscle fiber layer: is considered as a voltage source with an internal resistor R_M and a natural capacitor C_M of the muscle fiber.

In this model, we assume that values of resistors and capacitors are similar in each area [2, 9]. To obtain the graph of sEMG, a 2-D state space model can be utilized as shown below

$$\vec{X}' = A\vec{X} + B\vec{U} \tag{9}$$

where A, B are state space matrices [7], $\vec{U}$ is the input voltage vector, and $\vec{X}$ is the derivative of output voltage vector. In our work, the intracellular potentials of the fiber given by combining of equations (1) and (3) will be fed to the input voltage vector to create respective sEMG signals. These desired signals can be obtained by measuring output voltages across skin resistors. The simulated values of resistors and capacitors can be found in [2, 9].

2.4 Simulation Results of sEMG Signal

In order to obtain simulation results in time domain, the relationship in (3) will be used. Fig.4 presents simulation results of the intracellular potential and its derivatives in time domain. This potential changes quickly from -90mV up to +40mV and it includes three periods: threshold, depolarization and overshoot. A phase of the membrane named *After Hyperpolarization* which creates instantly a monopolar electrical burst is shown after 1.5ms. This figure also illustrates derivatives of the intracellular potential in which algebraic signs of the first derivative affect the orientation of the dipoles: depolarization with the positive sign and repolarization with the negative sign.

If the fiber diameter is changeable, it is well known to see effects on the intracellular potential as shown in Fig. 5. Here, we assume that the fiber diameter changes from 25µm up to 85µm [14]. With this change, minimum and

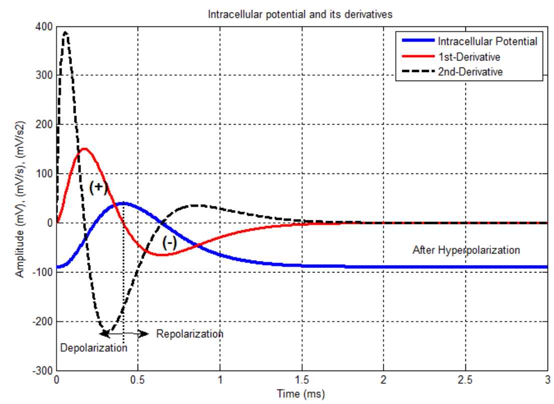


Fig. 4 Simulation results of the intracellular and its derivatives

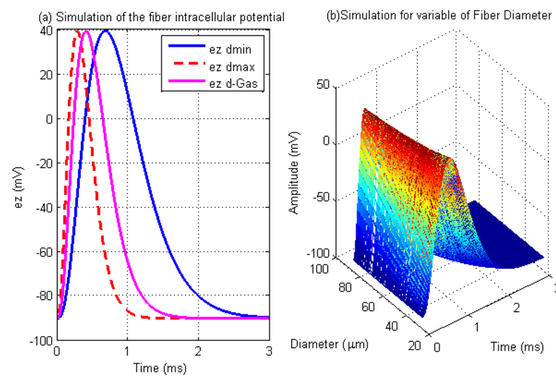


Fig. 5 Effect of fiber diameter on the intracellular potential when fiber diameter changes from dmin = 25µm up to dmax = 85µm.

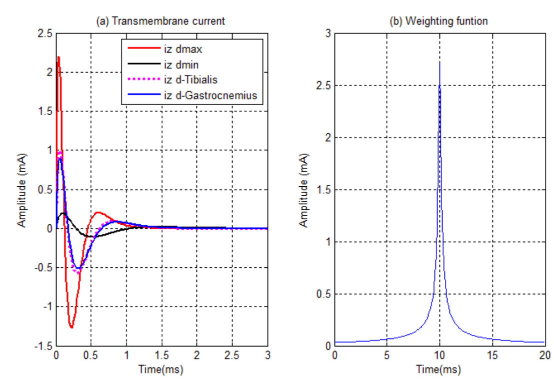


Fig. 6 Simulation of transmembrane current and weighting function. Simulation parameters are: dmax = 85µm, dmin = 25µm, d_Tibialis anterior = 57µm and d_gastrocnemius medialis = 54µm [15].

maximum values of the intracellular potential are not changeable, the corresponding times, however, are variable.

Figure 6 (a) shows the change of transmembrane currents corresponding to the variation of the fiber diameter which belongs to a square relationship. Figure 6 (b) gives the simulation result of a weighting function which is computed by [5].

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 1, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

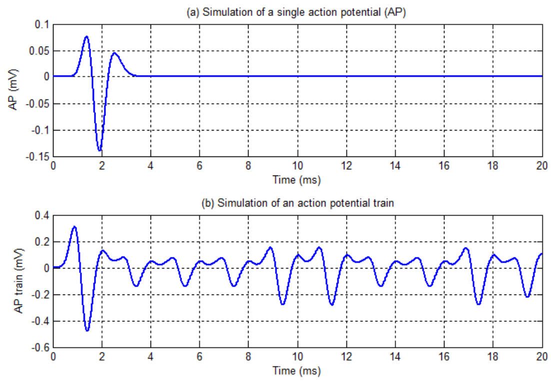


Fig. 7 Simulation of a sEMG signal train.

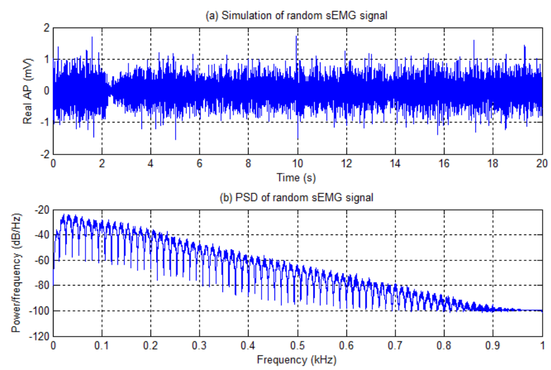


Fig. 8 Simulation of the random sEMG signal.

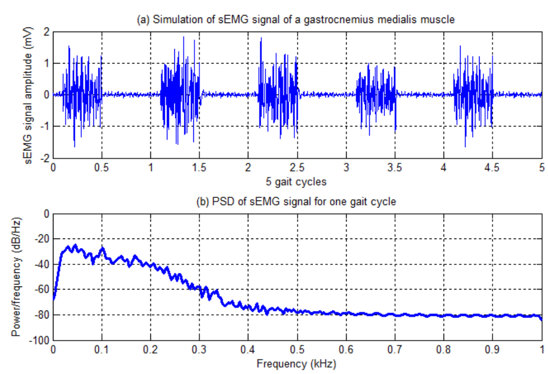


Fig. 9 Simulation results of sEMG signal of the gastrocnemius medialis muscle during gait cycles.

By using algorithms of FFT (*Fast Fourier Transform*) and iFFT (*inverse FFT*), we can compute and simulate the single fiber action potential as shown in Fig. 7. To achieve the superposed sEMG signal of a single fiber motor unit, we combine above motor units at different firing frequencies. In Fig. 7, sEMG signal of measured muscle is well known to reflect the recruitment and firing properties of the discovered motor unit.

According to the second sEMG signal model, simulation results are shown in Fig. 8. Here, conditions of simulation technique are used similarly with the above Gaussian sEMG signal (Fig. 2) to compare respectively simulated results. Their shapes are clearly close and appropriate to the actual sEMG signal.

With three above models of sEMG signal, aims of simulation are implemented successfully. Figure 9 illustrates the sEMG signal of gastrocnemius medialis muscle which is one of three important muscles in each leg during human walking with five gait cycles that can make a great significance for our exoskeleton system. Simulated parameters are found in [15] and this result is also appropriate for real experiments [15].

3. Analyzing of EMG-Based Control Schemes for LLPA Exoskeleton

In this part, we analyze several control strategies for LLPA exoskeleton based on EMG signal to approach our system. Figure 10 (a) describes a basic control scheme for our work. As shown in this diagram, measured EMG signal is fed to EMG signal processing block to make control signals [1]. Many methods have been investigated successfully to process EMG signal with high-pass filtering, rectification, and then low-pass filtering to obtain a control signal [1, 3, 6]. By combining with the other control blocks as muscle model, gain and controller, needful control signals are fed to LLPA exoskeleton system. These signals are normally used for torque and force of joints to control an exoskeleton robot [7, 10]. Feedback signals that can be used in this model are torque and position signals to implement feedback control for this system [3]. The advantages of this control scheme are simple to construct and implement. Its drawbacks are lower efficiencies and limited applications.

Recent researches have been improved significantly above disadvantages by giving new control schemes [3, 7]. In these systems, interaction between human and exoskeleton is concentrated more efficiently [3, 7, 10] and sensor systems are used for both robot and human as shown in Fig. 10 (b). In this control model, sensors are only used for the robot, such as encoders. Human sensors are applied to measure human parameters as angle, force, velocity, and acceleration of ankle or knee. These parameters are useful to control an exoskeleton for other goals of stroke patient rehabilitation [12] or extending the strength of humans [3]. These systems can be called as exoskeleton robot intelligent systems.

In order to apply a model for our work, besides above sensors, an open-circuit respirometer is used to measure oxygen consumption and carbon dioxide production to estimate human metabolic cost during walking with or without LLPA exoskeleton and with carrying military loads of 20 kg, 40 kg, and 55 kg. According to the aim of our exoskeleton system, the most important design tasks are to reduce the metabolic cost of locomotion and minimize the power requirements of assisted robot. With these tasks, sEMG signal-based control is able to achieve lower metabolic cost than those under kinematic control.

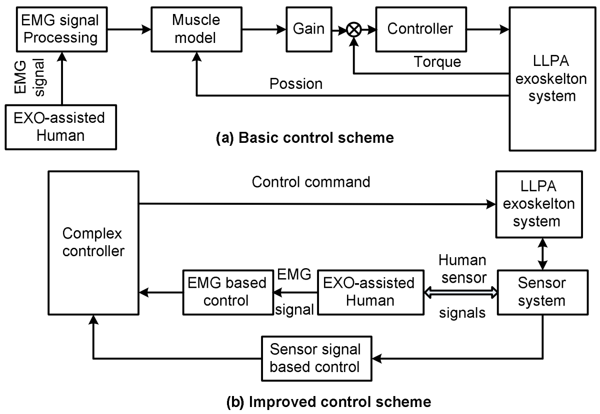


Fig. 10 Diagram of EMG-based control schemes.

4. Conclusions

In this paper, sEMG signal simulation techniques and control schemes of EMG-based LLPA exoskeleton system are presented. Although simulation techniques of sEMG signal are vast, three models of them are discussed as the most popular approaches. The first is the most basic, the second is the most easy, and the third is the most exact to simulate sEMG signal for goals of study and evaluation of simulated models with real systems. The other part of this paper explains two general models of EMG-based control algorithm for a LLPA exoskeleton in which the second one is applied for our work as an exoskeleton robot intelligent system. In the near future, we will perform this model and compare with real system to reach successfully given tasks in the exoskeleton project in our laboratory.

Acknowledgments

The authors would like to thank members of PRMI laboratory for assistance to implement this exoskeleton project.

References

[1] A. Salman, J. Iqbal, U. Izhar, U.S. Khan, and N. Rashid, "Optimized circuit for EMG signal processing", *2012 International Conference on.* Robotics and Artificial Intelligence (ICRAI), 2012. 208-213.

[2] A.K. Mahabalagini, K. Ahmed and F.Schlereth, "A novel approach for simulation, measurement and rpresentation of surface EMG signals", *Conference Record of the Forty Fifth Asilomar Conference on.* ASILOMAR 2011: Signals, Systems and Computers, 2011. 476-480.

[3] Ram Murat Singh and S. Chatterji, "Trends and challenges in EMG based control scheme of exoskeleton robots - a review." *International Journal of Scientific and Engineering Research,* 3(8), 2012: 1-8.

[4] Gregory S. Sawicki and Daniel P. Ferris, "Mechanics and energetics of incline walking with robotic ankle exoskeletons", *Journal of Experimental Biology*, 212(1), 2009: 32-41.

[5] J.Duchene & J.-Y.Hogrel, "A model of EMG generation", *Biomedical Engineering, IEEE Transaction on,* 47(2), 2000: 192-201.

[6] Kazuo Kiguchi and Y. Imada, "EMG-based control for lower-limb power-assist exoskeletons", *IEEE Workshop on.* RIISS '09: Robotic Intelligence in Informationally Structured Space, 2009. 19-24.

[7] H.He and K. Kiguchi, "A study on EMG-based control of exoskeleton robots for human lower-limb motion assist", *6th International Special Topic Conference on.* ITAB 2007: Information Technology Application in Biomedicine, 2007. 292-295.

[8] Konral, Peter. *The ABC of EMG.* A Practical Introduction to Kinesiological Electromyography, 2005.

[9] Fansan Zhu and N. W. Levin, "An electrical resistivity model of segmental body composition using bioimpedance analysis", *Proceedings of the 25th Annual International Conference of the IEEE.* Engineering in Medicine and Biology Society, 2003, 2003. 2679-2682.

[10] D.P. Ferris and C.L. Lewis, "Robotic lower limb exoskeleton using proportional myoelectric control", *Annual International Conference of the IEEE.* Engineering in Medicine and Biology Society, 2009. 2119-2124.

[11] Dario Farina and Roberto. Merletti, "A novel approach for precise simulation of the EMG signal detected by surface electrodes", *Biomedical Engineering, IEEE Transactions on*, 2001: 637-646.

[12] O. Unluhisarcikli, M. Pietrusinski, B. Weinberg, P. Bonato, and C. Mavroidis, "Design and control of a robotic lower extremity exoskeleton for gait rehabilitation", *2011 IEEE/RSJ International Conference on.* Intelligent Robots and Systems (IROS), 2011. 4893-4898.

[13] S. Patrick, J. Meklenburg, S. Jung, Y. Mendelson, and E.A. Clancy, "An electromyogram simulator for myoelectric prosthesis testing", *Proceedings of the 2010 IEEE 36th Annual Northeast.* Bioengineering Conference, 2010.

[14] Sanjeev D.Nandedkar, E.V. Stalberg and Donald B.Sanders, "Simulation techiques in electromyography", *Biomedical Engineering, IEEE Transaction on* 32, no. 10 (1985): 775-785.

[15] W. Wang, A. De. Stefano, and R.Allen, "A simulation model of the surface EMG signal for analysis of muscle activity during the gait cycle", *Computer in Biology and Medicine,* 36(6), 2006: 601-618.

Ngoc Khoat Nguyen received Msc degree in *Automation and Control* at Hanoi University of Science and Technology, in 2009. He is working as a Lecturer at Electric Power University in Hanoi, Vietnam. He is now doing PhD in the field of Automation at University of Electronic Science and Technology of China, Chengdu, Sichuan, China. His research interests include modeling and application of sEMG signal, control motor and robot.

Thi Mai Phuong Dao received Msc degree in *Automation and Control* at Hanoi University of Science and Technology, in 2010. She is working as a Lecturer at Hanoi University of Industry in Hanoi, Vietnam. She is now doing PhD in the field of Automation at Hunan University, Changsha, Hunan, China. Her research interests include modeling and experiment of sEMG signal, control theory and application.

Supporting Context-Centric Relations in Heterogeneous Environments

Jamie Walters, Theo Kanter and Rahim Rahmani

Department of Computer and System Sciences, Stockholm University
Forum 100, Kista, Sweden

Abstract
Massive Immersive Participation is enriched through the use of context information describing the dynamic states and relations among people places and things. This in turn mandates the creation of methods and models for establishing and supporting these relationships. Previous approaches are undermined by their limited interpretation of context centric relations and subsequently do not offer support for multi-criteria relationships. In this paper, we extend on our previous work on establishing multi-criteria context relationships, to adding the support required for maintaining these relationships over heterogeneous and dynamic context information. We introduce a query language that supports an extended publish-subscribe approach and define solutions for dynamically evaluating and adjusting these relationships while minimizing overall costs.

Keywords: *Context-Awareness, Immersive Participation, Context, Context Models, Internet Of Things, Context Proximity, Sensor Information, P2P Context*

1. Introduction

The increase in massive immersive participation scenarios is one part of our trend towards a pervasive computing reality. Such realities range from immersive games such as Google Ingress [1] to theatrical productions such Maryam [2] produced by RATS Theatre [3]. Users are immersed in realities that range from world domination to complete theatrical performances where people places and things are fused together in dynamic participatory environments.

The resulting immersion is enriched through the addition of the underlying context information driving the interactions among the collection of connected things. Supported by an Internet of Things (IoT), this additional information presents itself as the backbone of our pervasive realities, which responds to and accommodates for the establishing of the dynamic relationships that exist between a user, his environment and services. Systems such as SenseWeb [4], IP MultiMedia Subsystem (IMS) [5] , MediaSense [6] and SCOPE [7] were developed in response to this need to provision information supporting immersive realities.

Their limitations with respects to expressivity however limits their suitability in answering the question of *"Who you are, who you are with and what resources are nearby"* as required by Schilit and Adams suggest in [8] and further reiterated and summarized by Dey in [9], who expects that applications and services be provided answers to the question of *[which] entity is considered relevant to the interaction between a user and an application*.

While semantic approaches such that described by Dobslaw et al. in [10], Toninelli et al in [11] and Liu et al. in [12] offer some support towards this problem, Adomavicius et al. in [13] suggested that these types of approaches are limited and should be complemented by metric type approaches thus realizing the ability to answer the question of *"nearness"* as posited by both Schilit et al. in [8] and Dey and [9]. This further characterization would permit us to better identify and establish context relations between related entities. Therefore, establishing the types of relationships shown in Figure 1 is premised on our ability to support the complementing metric-type similarity models which, according to Hong et al. in [14], is critical in realizing applications and services that can discover nearby sensors or points of information.

Supporting these types of relationships is a multifaceted problem involving the identification and selection of candidate entities and managing the subsequent volume of required context information. One approach to this is through the use of centralized presence systems such as described by Petras et al. in [15]. Here an entity watches other entities contained in its address book. While this reduces the volume of context information required to maintain relationships, the resulting relationships are not context centric and limits the *watcher's* ability to discover entities of interest with which to establish common context relationships. With the average address book estimated to be limited in size to *0.005 * population* [15], the alternative of subscribing to all users would not present itself as a feasible solution with the volume of messages per status change would be approximated to *population x population*. This solution would not scale well and simply pruning the message queue as suggested by Petras et al. would offer little guarantee with regards to the quality of the context information.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 1, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

In defining an *Operational Approach to Context* Zimmermann et al. in [16] describes the notion of proximity as the overarching factor in establishing context relationships. This subsumed the earlier address book approaches and moved towards realizing truly context-centric networks where interactions, discovery and relationships are underpinned by the degrees of relationships between entities over their underlying context information. Zimmermann et al. equated the notion of proximity to spatiality, essentially disregarding the types of higher level relational proximity expected by Hong et al. in [14].

In [17] we defined an approach to establishing context centric relationships between entities on an Internet of Things. Here, relationships are established between entities over the similarity of their underlying context behaviors evaluated over a pre-determined time window. This extends the work of Zimmermann et al. in [16] towards a context-centric model while subsuming it with respects to expressiveness. This satisfies the initial requirement of a context relational model capable of supporting the establishing, adjusting and exploiting of implicit context based relationships in massive immersive environments. With this approach, we are capable of identifying and discovering candidate entities that can be fused to realize new user experiences and deliver more expressive applications and services.

However the problem of support remains as Zimmermann et al. in [16] provided no solution for discovering the candidate entities and establishing relationships in light of the highly dynamic nature of context information. Schmohl partially addressed this in [18] proposing a multi-dimensional hypersphere of interest in which entities entering are deemed to be candidates for the *watcher* and are evaluated and selected according to a proposed proximity measure. Here entities are discovered through the use of multi-dimensional indexing structures such as R-trees, kd-trees and space partitioning grids. These solutions are less optimal for multi-dimensional dynamic context environments as the cost of indexing increases exponentially with a linear increase in the number of dimensions. Queries therefore risk being executed against outdated indexes with no guarantees of information freshness. As a solution to this problem, Schmohl suggested that dimensions could be selectively pruned from the indices. By taking this solution applications depending on less popular dimensions would not stand to realize any benefit from this optimization. Alternatively, accuracy could be sacrificed for speed, which would not offer any guarantee of information accuracy for applications where this is critical.

Yoo et. al in [19] and Santa et. al in [20] proposed the use of publish-subscribe approaches as suitable alternatives with Kanter et. al in [21] showing that such approaches are scalable and can realize dissemination times on par with UDP signally used in SIP implementations. Frey and Roman in [22] extended this approach to provide for event driven subscriptions in context networks, however this is based on events rather than the raw underpinning context information.

Supporting the establishing of context-centric relationships over heterogeneous context information therefore requires new approaches to maximizing the identification of candidate entities while minimizing the overall resource costs

In this paper we introduce an approach to discovering related context entities through the use of an extended publish subscribe module coupled with a context query language. Additionally we introduce solutions for reducing the corresponding resource demands on established relationships. In Section 2 we summarize our approach to establishing context relationships. Section 3 details our approach to supporting these relationships while Section 4 discusses early analysis and results. Section 5 completes with a conclusion and discussion.

2. The Context Relational Model

In [17], we introduced a dynamic heterogeneous approach to context relationships where notion of context proximity is one that considers the situation, attributes, relations, accuracy and heterogeneity of both the underlying information and the vast array of requirements for metrics supporting application domains. We defined context proximity as: *the amount of work required to transform the context behavior of one entity into that of over the characteristics of their current underlying context states*

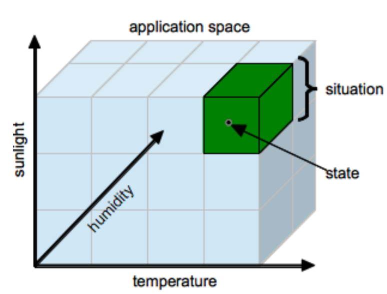


Figure 1 - Context Relational Model

Here, we model context as an n-dimensional domain space; the universe of discourse of a problem domain; the subset of all global information considered relevant and supports the delivery of any application or service relative to this domain.

In an immersive participation environment, such a domain could be the play "*Maryam*". This domain is then partitioned into situations or activities representing an acceptable range of context information defining a real world situation or activity. For the domain *Maryam*, activities could be *Scene 1, Scene 2, Scene 3*, etc. Activity definitions are not mutually exclusive and therefore several activity sub-spaces can overlap.

Each situation in turn contains context states; a combination of unique attribute values within a situation or activity space corresponding to a context observation made on an entity. For the domain *Maryam,* a state could be the context information recorded from body sensors at *Scene 1*.

2.1 Activity Classification

Citing the lack of consideration given to the context activity by existing proximity approaches, an underpinning element of our approach is use of activity information for deriving relational proximity. We identify activities of using the probabilistic approach described by Padovitz in [23].

The activity of an entity P bearing a context state $\forall S^{\mathcal{A}} = \{s_1, s_2, \dots\ s_i\} \vdots\ s_i \in \mathcal{A}_i$ can be determined by assigning it the activity with the highest confidence calculated as:

$$q_1 \sum_{i=1}^{n} \widehat{w}_1 \,.\, \Pr(a_i^t \in A_i) + q_2 \prod_{k=1}^{m} Pr(a_k^t \in A_k) \qquad 1$$

where $q1 + q2 = 1$. This is discussed in detail in [24]. With this approach, we consider state value membership, information accuracy and the importance of each context attribute to determining an activity. Here, we could observe states in *Maryam* and identify the current activity being experienced by the user.

2.2 Activity Similarity

We define a similarity matrix between the activities within an application or domain space.

	stand	walk	sit	lay
stand	1	0.7	0.5	0.25
walk	0.7	1	0.4	0.05
sit	0.5	0.4	1	0.70
lay	0.25	0.05	0.70	1

Figure 2 Activity Similarities

As shown in Table 1, this is an MxN matrix of real values between 0.0 and 1.0 conveying the similarity between activities in a domain space; the ease with which one activity can be transformed into another. Here an activity is not simply confined to primitive determinations such as walking, running, sitting laying but rather encompasses higher level notions of activities such as *going to work*, *going home*, *shopping*, *watching television*, *washing*, *cooking*.

These higher-level activities are not necessarily discernable from raw context information but can be derived by applying learning methods, human annotation and assumptions. The underlying context information could be very similar or even identical while the perceived higher-level activities are not.

2.3 Relational Context Proximity

Relational proximity is derived between the states of entities as observed over a time window W. For solving this, we used the Earth Movers Distance as described by Rubner et al. in [25] setting the distributions as the sets of observable context states for each window W, the weighted edges being the activity similarity between P and Q and the ground distance d_{ij} being the distance between pairs of states s_i, s_j derivable as:

$$d_{ij}(s_i, s_j) = \frac{\left(\sum_{k=1}^{n}\left[w_a * \left|\mathcal{F}_a^{\mathcal{D}}(a_i, a_j)\right|^r\right]_k\right)^{\frac{1}{r}}}{\left(\sum_{k=1}^{n}\left[w_a * \left|\mathcal{F}_a^{\mathcal{D}}(a_i, a_j)_{max}\right|^r\right]_k\right)^{\frac{1}{r}}} \qquad 2$$

where $a_i \in \mathcal{A}_i^{\mathcal{D}}, a_j \in \mathcal{A}_j^{\mathcal{D}}$

Here, w is the weighting for each attribute. The value of r can be adjusted to reflect the perceived distance between P and Q as shown by Shahid et al. in [26]. The distance is normalized with respects to the maximum distance between states in the encompassing application space. Our measure of proximity therefore logically subsumes and extends existing $Lp - norm$ approaches.

The EMD algorithm is then applied to derive the largest possible transformation between P and Q that minimizes the overall context transformation cost, where:

$$WORK(P \to Q, \boldsymbol{F}) = \sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} d_{ij} \qquad 3$$

Subjected to the following constraints:

$$\begin{aligned} &1.\quad f_{ij} \geq 0 \quad 1 \leq i \leq m,\ 1 \leq j \leq n \\ &2.\quad \textstyle\sum_{i=1}^{m} f_{ij} \leq P \quad 1 \leq i \leq m \\ &3.\quad \textstyle\sum_{j=1}^{n} f_{ij} \leq Q \quad 1 \leq j \leq n \\ &4.\quad \textstyle\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} = min\left(\sum_{i=1}^{m} P, \sum_{j=1}^{n} Q\right) \end{aligned} \qquad 4$$

The first constraint permits the transformation and hence the proximity from $P \to Q$ and not the opposite. The second and third constraints limit the transformation $P \to Q$ to the maximum number of context observations made for P or Q. The final constraint forces the maximum transformation possible between both entities. The context proximity, $\delta_{(P,Q)}$, is the earthmover's distance normalized by the total flow.

$$\delta_{(P,Q)} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} d_{ij}\right) * \left(\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}\right)^{-1} \qquad 5$$

We using the maximum possible flow between P and Q. It is important to note, that $\delta_{(P,Q)}$ is indifferent to the size of both sets of observations and permits partial similarity where the behavior of P is subsumed by the behaviour of Q. Therefore $\delta_{(P,Q)} \mid w = \delta_{(P,Q)} \mid \frac{1}{2} w$. This is a distinct advantage of our approach and excess observations are inherently discarded.

However, where partial matching is desirable and the completeness of containment is important for relations such that $P \cap Q = P \cup Q$, we extend the proximity measure to be normalized relative to the maximum potential transformation of either P or Q, such that

$$\delta_{(P,Q)} = \left[\left(\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} d_{ij}\right) * \left(\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}\right)^{-1}\right] \quad 6$$

$$\left|\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} = max\left(\sum_{i=1}^{m} P, \sum_{j=1}^{n} Q\right)\right.$$

The Confidence Constraint

In order to consider scenarios over unreliable context information, we can adjust the distance d_{ij} to reflect the potential errors in the underlying context information such that:

$$d_{ij} = d_{ij} * \left[c_{ij} + \left((1 - c_{ij}) * (1 - k)\right)\right]$$

$$where\ c_i = \sum_{i=1}^{n} \widehat{w}_1 \,.\, \Pr(a_i^t \in A_i) \quad 7$$

This confidence measure is described by Padovitz et al. and considers the accuracy of the sensors using several factors described in [24]. However for scenarios where the confidence is a trade off, we, we add the confidence factor k, which allows us to adjust this trade-off.

The Temporal Constraint

For calculating proximity considering the time constraint, we can adjust the size of the observation window W. For clarity:

$$\lim_{W \to 1} EMD\,(P,Q) = Lp_{norm}(P,Q) \quad 8$$

By adjusting W we permit wider variations in the time differences between state observations reducing the time constraints. Increasing W increases the constraint on the nearness of observations with respects to their temporal attribute.

The Continuity Constraints

Furthermore, we derive the measure of proximity stability between two entities as a means of filtering entities based on the stability of potential relationships. The first constraint finds the standard deviation of $\delta_{(P,Q)}$ as the window W progresses. We call this the co-relational constraint defined as:

$$R_{(P,Q)} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(\delta_{(P,Q)_i} - \mu\right)^2} \quad 9$$

$$and\ \mu = \frac{1}{N}\sum_{i}^{N} \delta_{(P,Q)_i}$$

Where the greater the deviation the more unstable the relationship between is. Secondly, we derive the convergence factor between two entities; the rate at which their context proximity is converging defined as**:**

$$Cf_{(P,Q)} = \frac{\Delta\delta_{(P,Q)}}{\Delta W} \quad 10$$

With this factor we can consider entities that are diverging or moving apart or entities that are getting closer or merging over time. Having established context relationships of the types described we are subsequently required to maintain these relationships at a minimal cost overhead.

3. Supporting the Context Relational Model

In Section 2.3, we described an approach to establishing context relationships over a proximity function defining the relationship between the behaviors of two given entities. Supporting this type of relationship requires an approach to finding, establishing and maintaining the relationships between the entities satisfying the proximity function. Figure 3 illustrates this process of supporting relationships over relational proximity.

Firstly, we created a query language for defining the proximity; the bounds of the hypercube of interest. The query is then executed across a distributed heterogeneous data store with the candidate entities selected for establishing a relationship. The resulting relationships are maintained by subscribing to the entities of interest and continually evaluating the relationships with each derived context state ranking each entity by its current proximity. New entities are continually added while non-relevant entities are consequently pruned.

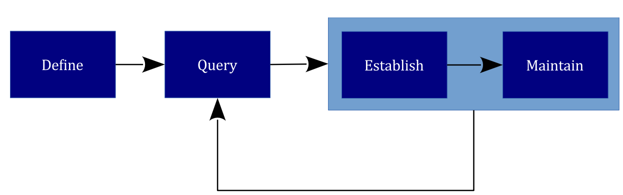


Figure 3 Supporting Relational Proximity

Finding related entities requires that the underpinning context information is readily stored and accessible and is organized in such that it permits discovery in a scalable manner with respects to real time qualities. While centralized approaches may be supported by databases, they are inhibited by their inability to scale well in response to service demands and availability. In response to this, we proposed a distributed approach to organizing context information using a self organizing peer-peer protocol [27].

3.1 Distributed Organization

At the base of our approach is the existence of a distributed data store capable of locating context information in response to the interest of applications and services. A fundamental point of departure from similar approaches is that this distributed data layer is capable of responding to queries on ranges of data and capable of answering queries with a range. Our solution is detailed in [27]and consists of an organized peer-to-peer data store with a lookup complexity of $0.5 \log n$, and a protocol for persisting, locating and subscribing to entities of interest. This permits us to find entities based on an area of interest by defining range of context values encompassing this area of interest. However defining areas of interest requires more expressive means of expressing context based queries than the primitive constructs of the associated protocol defined in [28].

3.2 Relational Context Query

In response to this we introduce a declarative query language, the Context Proximity Query Language (CPQL), for defining an area of interest relative to an entity. This is a natural extension of the interest based approach we introduced in [21], however with the interest area defined as a complex distance function over any underlying context information.
Similar to type and structure to SQL, the query language sits at the core of our query functionality and has two main constructs: *GET* or *SUB*. A *GET* is similar to an *SQL-SELECT* and retrieves all the states that currently match the defined proximity function. This can be used for a single evaluation for finding candidate entities. A *SUB* is also the equivalent to the *SQL-SELECT* with the addition that its function is to supporting existing established relationships between the entities bearing the satisfying states through an extension of publish-subscribe approach as described by Kanter et.al in [6]. Each application space defines function modules that satisfy $\mathcal{F}_a^{\mathcal{D}}(a_i, a_j)$ in Equation 2. Function modules are persisted on the distributed overlay architecture described in [6] and are identifiable, and retrievable by all nodes within the distributed data space.

A CPQL query is the defined of the type:

GET|SUB PRESENTITY
WHERE DISTANCE $DIST\_NAME$ $<$ 0.25
[ORDER ASC|DESC]
DEFINING DIST_NAME
AS sqrt(pow($\mathcal{F}_{lat}(P_{lat}, Q_{lat})$, 2) + pow($\mathcal{F}_{lon}(P_{lon}, Q_{lon})$, 2)); 11

For modularity and re-use, proximity functions can be defined prior to usage be saved as:

DEFINING DIST_NAME
AS sqrt(pow($\mathcal{F}_{lat}(P_{lat}, Q_{lat})$, 2) + pow($\mathcal{F}_{lon}(P_{lon}, Q_{lon})$, 2)); 12

Successive queries reference the defined function as:

GET|SUB PRESENTITY
WHERE DISTANCE $DIST\_NAME$ $<$ 0.25
[ORDER ASC|DESC] 13

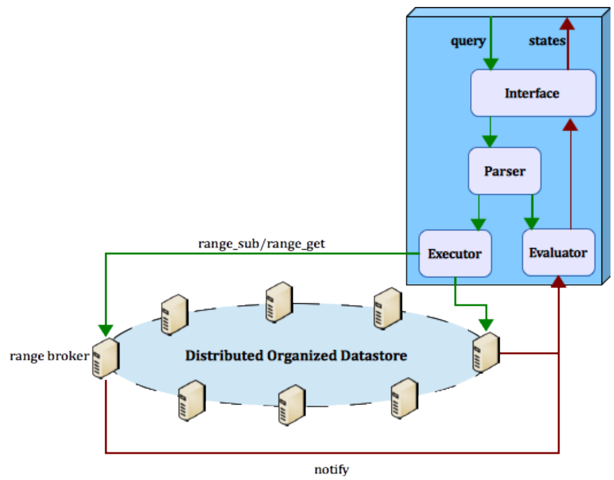


Figure 4 Relational Proximity Query-Subscribe Model

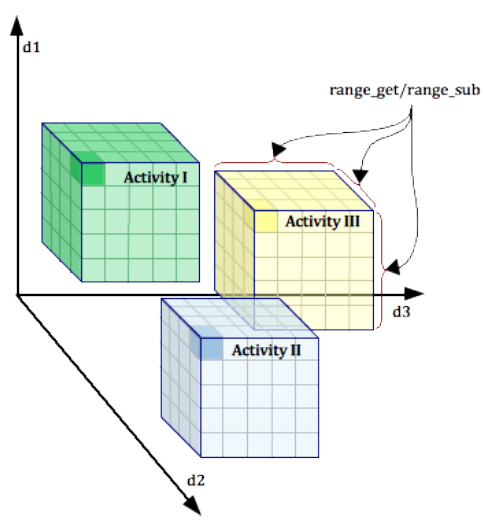


Figure 5 Range Query Subscribe

This permits us to share query definitions in a global data store that can be found and re-used by applications and services. Queries are accepted by an interface layer, which also provides the support for defining proximity definition functions. Each query is parsed and analyzed for correctness and decomposed into its corresponding parse tree. The resulting tree is passed to an executor for execution across the distributed data store. The proximity function as described in Equation 9 uses $(\boldsymbol{a_{min}}, \boldsymbol{a_{max}})$ as the upper and lower boundaries for each dimension of the application space and are used as the limits for the *range_sub* and *range_get* as shown in Figure 5. A range get for any proximity function is then expressed as: range_get$(\boldsymbol{a_{min}}, \boldsymbol{a_{max}})$. This limits the context states and entities that are queried to only those being relevant to and entity for the execution of a specific application or service.

3.3 Relational Publish Subscribe

Establishing relationships requires a new approach towards publish-subscribe solutions in context aware systems. Previous publish subscribe solutions such as that detailed by Kanter et. al in [6] supported primitives for getting or subscribing to the context information of an entity. Unlike Petras et al. in [15] such a solution was distributed reducing issues of scalability; however like Petras et al. a watcher cannot subscribe to greater than the size of its address book. This is estimated in [15] to be around 0.005*N,* the number of global presentities. However, in context centric approaches, a watcher's address book does not determine the number of presentity it watches. This should be determinable by the number of entities with which it can potentially establish a relationship. The absolute maximum number of subscriptions for each watcher would therefore be *N* which is not scalable.

To address this shortcoming and provide scalable support for context centric networks, we extend the publish-subscribe approaches to enable subscriptions to relationships and areas of interest as defined by an underlying proximity function. A *watcher* issues a subscription for all (a_{min}, a_{max}) of the underlying application space. This is issued to *range_brokers*, distributed nodes responsible for brokering ranges of values between *watchers* and their *presentities*. The *range_sub* or *range_get* is routed by the underlying support to the node or nodes in the data space shown in Figure 4 that are responsible for the range of values in (a_{min}, a_{max})**.** Each *presentity* each presentity subsequently publishes its current context value to the corresponding *range_broker*. The *range_broker* in turn sends the current set of states to the watcher. The *watcher* evaluates the list of states over its proximity function and establishes a context relationship with selected *presentities*. In order to

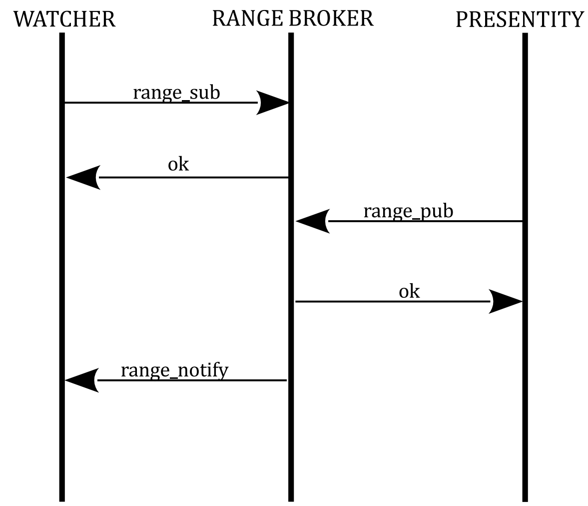


Figure 6 Range Publish-Subscribe Messaging

maintain the relationship, the *watcher* then subscribes to the *presentities* and continually evaluates the relationship with each new context states received. The *watcher* also maintains a subscription with the range*_broker*, which continually send lists of context states matching the range subscription to the watcher. Context entities publish new context states as the supporting context information changes; this can range from very frequently in highly dynamic situations to seldom in lesser dynamic situations. Petras et al. in [15] stated that the number of messages originating with each state change of a presentity is equal to the number of watchers for that presentity. In prescence systems without context information this can be taken as the size of the presentity's address book. However, in context centric approaches, the number of watchers for each presentity would have no relation to the size of such an address book.

Therefore, given that the known context universe contains *D* dimensions with *N* uniformly distributed entities. Each application *A* is enclosed by an application space with *M* dimensions such that $M \in \binom{D}{k}$ where $0 < k < D$. With existing publish-subscribe approaches the cost of subscribing to all entities is of the order $(N-1)$ for each entity and a global subscription cost of $(N^2 - N)$ for all known entities.

We however reduce the number of subscriptions, *n*, per entity by defining the number of potential candidate entities for each *watcher*. Given the universe of discourse of the application *A,* where each dimension *d* occurs with a probability distribution θ, the number of subscriptions, we limit the number of potential candidates to:

$$N \prod_{d \in M} P(d|\theta_d) \quad (14)$$

and the overall number of subscriptions for all entities contained in each application is:

$$\sum_{n \epsilon A} \left(N \prod_{d \in M} P(d|\theta_d) \right) \quad (15)$$

That is to say we do not allow subscriptions to entities without a probability of being a candidate entity i.e, entities without the required context dimensions. Additionally, given that an application space is further limited by the dimensions having values between (a_{min}, a_{max}), we further limit this to:

$$N \prod_{d \in M} P(d|\theta_d) * P(a_{min} < d_{val} < a_{max}|\theta_{dval}) \quad (16)$$

for each entity within the application space and:

$$\sum_{n \epsilon A} \left(N \prod_{d \in M} P(d|\theta_d) * P(a_{min} < d_{val} \right. \quad (17)$$

for all entities contained within each application. This is achieved through the use of the relational publish-subscribe approach. From this we derive the following proposition:

Proposition 1:
The number of presentities required for an application or service is limited by the universe of discourse of the application or service itself.

After selecting the candidate entities and establishing a relationship, we are now required to maximize the quality of the context relationship while minimizing the message overheads required for support. The maximum quality achievable by a context relationship is calculating with every change in the underlying context states. Current approaches to deriving context proximity between two entities calculates a notion of proximity either in response to continual changes in context information or on demand by the utilizing application or service. Therefore, the current state of the context relationship does not influence the flow of context information between two entities with an established relationship. From this we derive:

Proposition 2:
The cost of maintaining the context relationship between two given entities is determined by the state of the context relationship itself.

Given two entities P and Q, The cost of maintaining the relationship between any two entities $P \rightarrow Q$, is related to the current state. When the two entities are in close proximity the cost of maintaining their relationship is higher and decreases as the proximity between both entities approaches 1. Therefore with respects to proximity the $interval$ between refreshes of the underlying information is:

$$\begin{aligned} \lim_{\delta \to 0} interval\,(P,Q) &= interval_{min}(P,Q) \\ \lim_{\delta \to 1} interval\,(P,Q) &= interval_{max}(P,Q) \end{aligned} \quad (18)$$

We extend the publish-subscribe module to additionally accept a parameter for the *interval* delays between publishing messages to entities with an existing relationship. Entities, according to the model described in Section 2 generate states in response to changes in the underlying context information. Each change in the value of a context attribute generates a new context state. Each relationship has three values: i_{min} , the native interval between states for the entity Q, i_{min}, the maximum interval permitted by the application before the states must be refreshed and i_{cur} the current interval between context states calculated as:

$$i_{curr} = \begin{cases} i_{min} + \delta C f_{(P,Q)} * (i_{max} - i_{min}), \ , & f \neq 0 \\ i_{min} + \delta * (i_{max} - i_{min}), & f = 0 \end{cases} \quad (19)$$

The interval value is therefore calculated for each publish instance based on the last known proximity and the rate at which the proximity moves towards 0. This is an adaptive algorithm where the rate is kept relative to the known relationship quality. The intuition being that the closer two entities are the more resources that can be expended on maintaining their relationship, while distant entities require less updates and can make resources available to more critical relationships. The rate is adjusted based on the current rate of state generation, therefore we will not request updates faster than they are produced and not slower than the minimum required for the application's quality of experience. The closer the value of i_{min} is relative to i_{max}, the smaller the penalty. The intuition here being that, as $i_{min} \rightarrow i_{max}$, the back off potential gets smaller.

With the frequency of refreshes derived, we are now required to adjust the volume of context information used for maintaining the context relationships. Existing publish-subscribe modules transmit each updated element of context information to end points with no processing. Transmitting context information, which, in some cases offers no significant knowledge or variation in the perceived proximity between the entities. The context proximity model described in Section 3 creates a new state with every change in context information. The result being that context information such as GPS sensors, which might change continually would provide a continuous stream of information even in cases where the difference between states is marginal and does not serve to noticeably influence the resulting proximity values.

Firstly, we estimate the cohesion of points in I by randomly selecting a sample of the relationships between all the observed states. We take a sample set to avoid computing all the relationships within I, which is of order of complexity $O\,(n^2)$. The size of the sample set is determined by taking the normal approximation to the hypergeometric distribution for

$$n = \frac{N\,z^2pq}{(E^2(N-1)+z^2\,pq)}$$

n = *required sample size*
$N = I^2$ 20
p, q = *set to 0.5*
z = *confidence level, set to 1.96*
E = *accuracy, set to 0.03*
$0 < n < I$

This, as the hypergrometric approximation is more suitable for the relatively small numbers of states observed. We then sample n distances between states using the distance function described in Section 3.4, setting all the weights to 0.

With n selected distances, we then compute the cohesion as the standard deviation of the distances between states in I

$$cohesion = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}((s_1 - s_2)_i - \mu)^2} \quad 21$$

$$where\ \mu = \frac{1}{N}\sum_{i}^{N}(s_1 - s_2)_i\ and\ N = I^2$$

The final set of states transmitted for each observation window is therefore:

$$cohesion * I \quad 22$$

Here, the intuition is that a relatively stable entity expressing little change in its state over time would share less context information while an active entity would share almost all the context information generated. We therefore share as much information as is required for establishing and maintain the context relationships over time, deriving the following proposition:

Proposition 3:
The number of context states shared by an entity with each publish instance is a factor of the cohesion of the total set of observed context states over the observation interval

4. Evaluation

For simulation purposes, we assumed a global population N, of between 1000 and 100000 presentities each assigned a random context profile from a context universe D with 20 dimensions. The application A was assigned three random dimensions (d_a, d_b, d_c) such that $d \in D$.

The number of presentities with a dimension d was taken as a random value sampled from a binomial distribution $B\,(p/10, N)$ where the value of p is randomly sampled from the Gaussian distribution $\mathcal{N}\,(5.0, 1.6)$. The simulation was run 20 times for each network size and the results shown below in Figure 7. We show the min, avg and max for each network size.

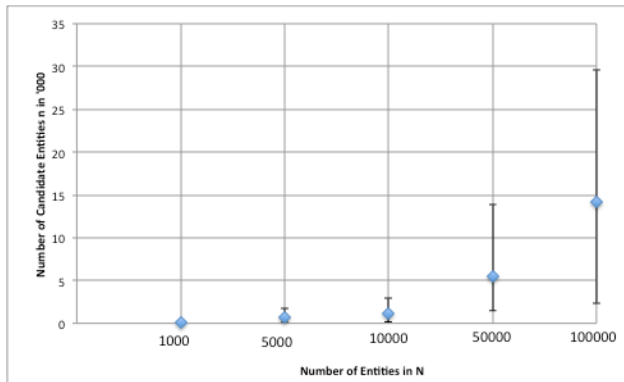


Figure 7 Candidate Entities

As can be seen from Figure 8 the number of potential subscriptions is reduced where subscriptions are made relative to the application's universe of discourse. The number of candidate entities vary widely for each simulation size demonstrating that a priori information such as address books cannot be used to determine the size number of subscriptions required by an application and is best supported by real time approaches such as publish-subscribe. As can be seen from Figure R.x the number of

N	Expected	Pub/Sub	Addressbook
1000	35	36	5
5000	177	186	25
10000	1485	1535	50
50000	2793	2700	250
100000	7611	7691	500

Table 1 Candidate Entities Comparison

Figure 8 shows the comparison between the number of subscriptions that would be issued for the publish subscribe and the addressbook approaches compared with the expected number of subscriptions based on the distribution of the underlying context dimensions. As shown, the publish-subscribe approach of locating entities based on their probability of being candidate maximizes the number of candidate entities while remains significantly less than a full search or subscription to every entity and avoids any expensive indexing approaches.

As shown in Figure 8, where entities are using applications and services with high duplication with respect to the attributes used for each application space, the number of subscriptions relative to applications increases at a lower rate. However, where there is little duplication among entities, the number of subscriptions quickly approaches N(100000).

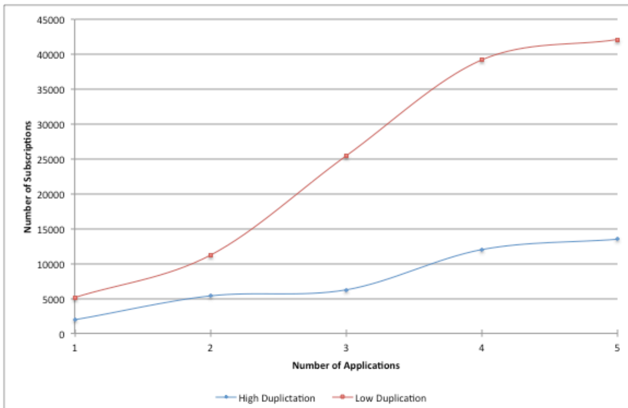


Figure 8 Subscription Increase With Applications

By applying the adaptive ate algorithm we further reduced the number of messages between entities as their proximity approached 1.0 and increased as the proximity approached 0.0. For this simulation we assumed that the distance between an entity is a normal distribution of proximity values between 0.0 and 1.0. Each entity updates its context information using a rate sampled from the Poisson distribution:

$$e^{-\lambda}\frac{\lambda^x}{x!}$$

Where λ is sampled from a set of uniformly distributed values between 1 and 30. This was simulated and the results shown in Table

Rate(msgs/min)	Adaptive Rate(msgs/min)
100184	29438
99799	29426
100159	29314
100052	29175

Table 2 - Adaptive Publish

The adaptive rate, reduced the frequency of updates for each period. This further reduces the overall messaging overhead needed to maintain the types of context relationships described in Section 2.

5. Conclusion

In this paper, we presented an approach to supporting context centric relationships between entities on an Internet of Things. This satisfies the requirement of solutions capable of supporting the types of dynamic relationships that exist over heterogeneous context information. With this approach, we are capable of identifying candidate entities that can be fused to realize new user experiences and deliver more expressive applications and services.

Firstly we proposed a context proximity query language for identifying candidate entities over a distributed data store supporting range queries. The queries are executed across a distributed heterogeneous data store with the candidate entities selected for establishing a relationship.
In order to address the shortcomings of previous approaches to context proximity and provide scalable support for context centric networks, we extended earlier publish-subscribe approaches to enable subscriptions to relationships and areas of interest as defined by the underlying proximity function. The resulting relationships are continually evaluated relationships with each derived context state, ranking each entity by its current proximity. As a result, new entities are continually added while non-relevant entities are consequently pruned. We further introduced an adaptive algorithm for adjusting the publish rate between two entities over the current state of their context relationship and adjust the volume of context information relative to cohesion of the observed context states.

We performed simulations to show that our approach allows us to maximize the number of candidate entities while reducing overall resource costs. We further showed that a priori solutions such as addressbook-based subscriptions do not provide sufficient support by subscribing to significantly less than the relevant number of entities. By avoiding high dimensional indexing and extending proven publish-subscribe approaches we benefit from its proven real time properties. Our solution scales well where applications are using common sets of context information, however where this is not the case, the number of subscriptions converges to N.

Future work includes the in network aggregation and evaluation of proximity as well as the derivation of activity similarity through crowd sourcing.

References

[1] "Ingress." [Online]. Available: http://www.ingress.com/. [Accessed: 26-Feb-2013].

[2] R. Forsberg and W. Sauter, "Digital theater in favor of audiences?," Stockholm, Sweden.

[3] "Kista Teater." [Online]. Available: http://www.kistateater.com/. [Accessed: 26-Feb-2013].

[4] A. Kansal, S. Nath, J. Liu, and F. Zhao, "Senseweb: An infrastructure for shared sensing," *IEEE MultiMedia*, vol. 14, no. 4, pp. 8–13, Oct. 2007.

[5] M. Strohbach, J. Vercher, and M. Bauer, "A case for IMS," *Vehicular Technology Magazine, IEEE*, vol. 4, no. 1, pp. 57–64, Mar. 2009.

[6] T. Kanter, P. Österberg, J. Walters, V. Kardeby, S. Forsström, and S. Pettersson, "The Mediasense Framework," in *2009 Fourth International Conference on Digital Telecommunications*, 2009, pp. 144–147.

[7] R. A. Baloch and N. Crespi, "Addressing context dependency using profile context in overlay networks," in *Consumer Communications and Networking Conference (CCNC), 2010 7th IEEE*, 2010, pp. 1–5.

[8] B. Schilit and N. Adams, "Context-aware computing applications," *Systems and Applications, 1994.*, pp. 85–90, 1994.

[9] A. K. Dey, "Understanding and Using Context," *Personal and Ubiquitous Computing*, vol. 5, no. 1, pp. 4–7, Feb. 2001.

[10] F. Dobslaw, A. Larsson, T. Kanter, and J. Walters, "An Object-Oriented Model in Support of Context-Aware Mobile Applications," *Mobile Wireless Middleware, Operating Systems, and Applications*, pp. 205–220, 2010.

[11] A. Toninelli, S. Pantsar-Syväniemi, P. Bellavista, and E. Ovaska, *Supporting context awareness in smart environments*. New York, New York, USA: ACM Press, 2009, p. 1.

[12] L. Liu, F. Lecue, N. Mehandjiev, and L. Xu, "Using Context Similarity for Service Recommendation," *2010 IEEE Fourth International Conference on Semantic Computing*, pp. 277–284, Sep. 2010.

[13] G. Adomavicius, R. Sankaranarayanan, S. Sen, and A. Tuzhilin, "Incorporating contextual information in recommender systems using a multidimensional approach," *ACM Transactions on Information Systems (TOIS)*, vol. 23, no. 1, pp. 103–145, 2005.

[14] J. I. Hong and J. Landay, "An infrastructure approach to context-aware computing," *Human-Computer Interaction*, vol. 16, no. 2, pp. 287–303, Dec. 2001.

[15] D. Petras, I. Baronak, and E. Chromy, "Presence Service in IMS.," *TheScientificWorldJournal*, vol. 2013, p. 606790, Jan. 2013.

[16] A. Zimmermann, A. Lorenz, and R. Oppermann, "An operational definition of context," in *Proceedings of the 6th international and interdisciplinary conference on Modeling and using context*, 2007, pp. 558–571.

[17] J. Walters, T. Kanter, and R. Rahmani, "Establishing Multi-Criteria Context Relations Supporting Ubiquitous Immersive Participation," *International Journal*, vol. 4, no. 2, pp. 59–78, 2013.

[18] R. Schmohl, "The Contextual Map," *deposit.ddb.de*, 2010.

[19] S. Yoo, J. H. Son, and M. H. Kim, "A scalable publish/subscribe system for large mobile ad hoc networks," *Journal of Systems and Software*, vol. 82, no. 7, pp. 1152–1162, Jul. 2009.

[20] J. Santa and A. F. Gomez-Skarmeta, "Sharing Context-Aware Road and Safety Information," *IEEE Pervasive Computing*, vol. 8, no. 3, pp. 58–65, Jul. 2009.

[21] S. Forsström, V. Kardeby, J. Walters, and T. Kanter, "Location-Based Ubiquitous Context Exchange in Mobile Environments," *Mobile Networks and Management*, pp. 177–187, 2011.

[22] D. Frey and G. C. Roman, "Context-aware publish subscribe in mobile ad hoc networks," in *Coordination Models and Languages*, 2007, vol. 4467, pp. 37–55.

[23] A. Padovitz, S. W. Loke, and A. Zaslavsky, "Towards a theory of context spaces," in *Pervasive Computing and Communications Workshops, 2004. Proceedings of the Second IEEE Annual Conference on*, 2004, no. March, pp. 38–42.

[24] A. Padovitz, "Context Management and Reasoning about Situations in Pervasive Computing," Monash University, 2006.

[25] Y. Rubner, C. Tomasi, and L. J. Guibas, "A Metric for Distributions with Applications to Image Databases," p. 59, Jan. 1998.

[26] R. Shahid, S. Bertazzon, M. L. Knudtson, and W. a Ghali, "Comparison of distance measures in spatial analytical modeling for health service planning.," *BMC health services research*, vol. 9, p. 200, Jan. 2009.

[27] J. Walters, T. Kanter, and E. Savioli, "A Distributed Framework for Organizing an Internet of Things," in *The 3rd International ICST Conference on Mobile Lightweight Wireless Systems*, 2011, pp. 1–17.

[28] O. Akan, P. Bellavista, J. Cao, F. Dressler, D. Ferrari, M. Gerla, H. Kobayashi, S. Palazzo, S. Sahni, X. (Sherman) Shen, M. Stan, J. Xiaohua, A. Zomaya, G. Coulson, K. Pentikousis, O. Blume, R. Agüero Calvo, S. Papavassiliou, S. Forsström, V. Kardeby, J. Walters, R. Norling, and T. Kanter, *Mobile Networks and Management*, vol. 32. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 57–66–66.

Extract an Essential Skeleton of a Character as a Graph from a Character Image

Kazuhisa Fujita

University of Electro-Communications
1-5-1 Chofugaoka, Chofu, Tokyo, 182-8585 Japan

Abstract

This paper aims to make a graph representing an essential skeleton of a character from an image that includes a machine printed or a handwritten character using the growing neural gas (GNG) method and the relative neighborhood graph (RNG) algorithm. The visual system in our brain can recognize printed characters and handwritten characters easily, robustly, and precisely. How can our brains robustly recognize characters? In the visual processing in our brain, essential features of an object will be used for recognition. The essential features are crosses, corners, junctions and so on. These features may be useful for character recognition by a computer. However, extraction of the features is difficult. If the skeleton of a character is represented as a graph, the features can be more easily extracted. To extract the skeleton of a character as a graph from a character image, we used the GNG method and the RNG algorithm. We achieved to extract skeleton graphs from images including distorted, noisy, and handwritten characters.

Keywords: *Skeletonization, Character Recognition, Self Organizing Map.*

1. Introduction

Why can we robustly recognize characters from a rotated, a distorted, and a noisy image including characters? This ability is provided by robust visual recognition mechanism in the brain. In the visual processing in the brain, common essential features of an object are used for recognition. Essential features of an object are crosses, corners, junctions, circles, and so on [5]. In pattern recognition by a computer, we may achieve to provide more robust image recognition if we effectively use the common essential features. The purpose of this study is to extract essential structures of a character from an image as a graph, that here we called a skeleton graph, in order to easily use essential structures of a character for character recognition by a computer. A skeleton graph represents the skeleton of a character. Each skeleton graph extracted from the images including same characters will be similar. Thus, using similarity of skeleton graphs allows us to achieve to develop more robust character recognition system. In this study, we propose the method of extraction of a skeleton graph of a character from an image including a character in order to develop more robust character recognition system directly using similarity of structures of skeletons.

Extraction of a skeleton from a character image is called skeletonization. Skeletonization is generally executed before recognition process by a learning machine [8]. Skeletonization is a general morphological method that is used to thin a broad stroke of a character image and to extract only a bone of a character from a character image. The major functions of skeletonization in image processing are to reduce data size and to make more easily extract morphological features. This conventional method allows us to extract a skeleton “image” of a character from an image.

To achieve to extract a skeleton graph, we employed the growing neural gas (GNG) method that is topology learning algorithm and one of self organizing map (SOM) methods. SOM [9] can be developed based on topology conserving classifiers. However, the network structure of SOM is static (generally, n-dimensional lattice) and the network structure cannot represent topology of input space. The GNG method improves this problem because the GNG method allows to flexibly increase or to flexibly decrease nodes and edges of the network. The GNG method has been proposed by Fritzke [4]. The GNG method has been widely applied to clustering or topology learning, such as reconstruction of 3D models [7], landmark extraction [2], and object tracking [3]. We applied the GNG method to skeletonization.

In the present study, we demonstrated making a skeleton graph of a character using our proposed method. Under noisy circumstance, our approach could also produce satisfactory result. This achievement may allow us to robustly extract an essential skeleton from a character images.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 1, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

2. Methods

2.1 Scheme

To make the skeleton graph, we used three steps. The first step, execute image processing that consists of binarizing and trimming. The second step, roughly extract a skeleton graph from a character image using the GNG method. The third step, remove redundant edges and rewire nodes using relative neighborhood graph (RNG) algorithm [12]. The GNG method is high ability to extract topological features of characters and easily method to assemble. However, the graph generated by the GNG method has a few redundant edges because the GNG method tends to make triangle clusters [1,6]. To resolve this problem, we used the RNG method that has the ability to extract a perceptually meaningful structure. Using this method, redundant edges reduce and only essential structures are extracted from a character image. Through the three steps, the skeleton graph can represent essential structures of a character.

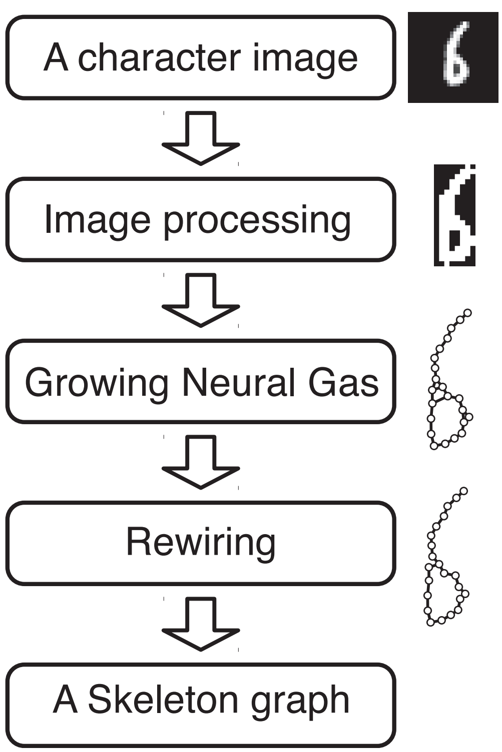


Fig. 1 The scheme to make a skeleton graph.

2.2 Growing Neural Gas

The growing neural gas (GNG) method has been proposed by Fritzke [4]. The GNG method is one kind of SOM methods and extracts topology or classifies data. The network of the GNG flexibly varies and its structure represents data structure. Using these features, we extracted the skeleton of a character from a character image as a graph.

The input space of the GNG network is an image that is a two-dimensional pixel space sized $W \times H$. The network consists of a set A of nodes. Each node $c \in A$ has an associated reference vector w_c. Node's reference vector w_c is denoted by $\boldsymbol{w}_c = (w_{cx}, w_{cy})$. The two parameters represent the node position over the image. The reference vectors must fulfill:

$$0 \leq \boldsymbol{w}_{cx} < W, 0 \leq \boldsymbol{w}_{cy} < H. \qquad (1)$$

There are edges between pairs of nodes. These connections are not weighted and not directed. The edges defined topological structure of the network.

Every node is examined to calculate which one's reference vectors are most like the input vector through the following process.

1. Starting with only two nodes that are connected each other. Positions of the nodes are random in $\mathcal{R}$.
2. The input vector $\boldsymbol{x}_i = (x_i, y_i)$ is chosen at random from pixels on a character.
3. The criterion for neighborhood is Euclidean distance between an input vector and a reference vector of a node. The number k of the winning (nearest) node is defined by,

$$k = \arg\min_i \ \|\boldsymbol{w}_i - \boldsymbol{x}\|. \qquad (2)$$

   Simultaneously, find the second nearest node s.
4. Increase the age of all the edges connecting with the wining node.
5. Add the squared distance between the input vector and the reference vector of the winning node to a local counter variable:

$$\Delta\text{error}_k = \|\boldsymbol{w}_k - \boldsymbol{x}\|^2 . \qquad (3)$$

6. The winning node k is rewarded with becoming more like the input vector.

$$\boldsymbol{w}_k(t+1) = \boldsymbol{w}_k(t) + \lambda(t)(\boldsymbol{x} - \boldsymbol{w}_k). \qquad (4)$$

   All direct neighbors n of k are also rewarded.

$$\boldsymbol{w}_n(t+1) = \boldsymbol{w}_n(t) + \lambda(t)(\boldsymbol{x} - \boldsymbol{w}_n), \qquad (5)$$

   where t is the learning frequency and λ is the learning coefficient. λ decays with the learning frequency.

$$\lambda(t) = \lambda_0 \times \left(1 - \frac{t}{T}\right), \quad (6)$$

where λ_0 is the initial learning coefficient and T is the preset maximum training step.

7. If k and s are connected, set the age of this edge to zero. If k and s are not connected, add the edge between these nodes.
8. Remove the edges with the age that is larger than $a_{\max}$. If the node isolated by this remove process, remove the node.
9. Every certain number of input signals generated, insert a new node. The number of nodes has limit $N_{\max}$.
   (a) Determine the neuron q with the maximum summed error.
   (b) If the maximum summed error is more than Error_0, insert new node r between q and its farthest neighbor f:

$$\boldsymbol{w}_r = (\boldsymbol{w}_q + \boldsymbol{w}_f)/2. \quad (7)$$

   (c) The values of Error_0 and $N_{\max}$ are required to hardly make redundant nodes and edges, and triangle cycles.
10. Every certain number of input signals generated, set all error variables to zero.
11. If a stopping criterion is not fulfilled, go to step 2.

Figure 2 shows the growing process of the skeleton graph generated by the GNG method. The character image includes "A". It can be seen that the GNG network learned the skeleton topology of the character.

The parameters for this simulation were: $\lambda = 0.2, N_{\max} = 40, a_{\max} = 28$.

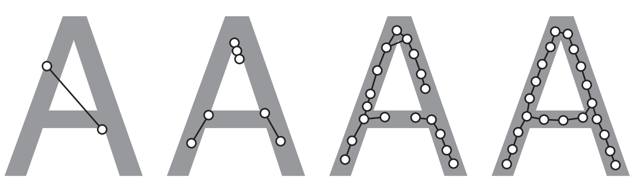


Fig.2 Different steps of convergence of the network for the character "A". These figures show the networks after 0, 4000, 12000, and 80000 steps (from left to right). At the end of the adaptation process the connection between the nodes represents the structure of "A".

2.3 Rewiring

The graph generated by the GNG method tends to have triangle cycles [1,6]. The skeleton graph generated by only the GNG method, shown in fig. 3 B, has the triangle cycles and the redundant edges. The triangle cycles and the redundant edges especially appeared on which stroke was crossed and on a broad curve line. To represent essential bone as a graph, deleting the triangle cycles and redundant edges of the skeleton graph generated by the GNG method are required. To reduce the redundant edges, we implemented the rewiring process. We kept nonredundant edges and deleted redundant edges using the Relative Neighborhood Graph (RNG) algorithm [11,12]. In the view of the RNG algorithm, each node of the graph is relative neighbor if they are near. If nodes i and j are relative neighbors, there dose not exist another node z of the set such that,

$$d(z,i) < d(i,j) \text{ and } d(z,j) < d(i,j), \quad (8)$$

where $d(i,j)$ is the Euclidean distance between i and j. When nodes i, j fulfill the equation except $d(i,j) > (w^2 + H^2)^{1/2} \times 0.15$, nodes are connected. Figure 3 C shows the skeleton graph processed by the RNG algorithm. Redundant edges reduced and an essential skeleton graph was extracted.

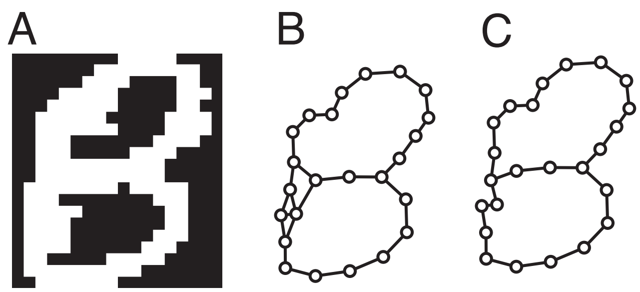


Fig. 3 A shows the handwritten digit "8". B and C illustrate the graph generated by the GNG method without and with the rewired process, respectively.

3. Results

To verify that the extracted skeleton graph represents essential structures of a character, the proposed method has been tested with the four sets of characters that are represented by binary images. The results for the four sets of images are shown in fig. 4. The first set consisted of regular images that include undistorted printed-characters. The skeleton graphs represented the essential structures of the characters, shown in fig. 4A. The second set consisted of the images of distorted and rotated printed-characters. In this case, the skeleton graphs also represented essential structures. The topology of the skeleton graphs is almost same structures of the skeleton graph generated from the regular images. The third set consisted of the images of isolated handwritten digits from the MNIST Database [10]. In this case, the skeleton graphs also represented essential structures.

The fourth set consisted of the images of noised characters. In this test, we randomly changed white pixels on a character to black pixels. The random noise is uniformly distributed on a character. Here, we define noise rate $\xi = \nu/\rho$, where ν is the amount of changed pixels and ρ is the original number of pixels on a character. Figure 4D shows the skeleton graphs at different noise level. The skeleton graphs produced by our method was consistent with visual form of characters for $\xi = 0.95$ and 0.99. However, for $\xi = 0.995$, the skeleton graph could not represent the form of characters.

Using our method, we could extract essential structures of a character as a skeleton graph. It is important that the skeleton graphs made from images including a same character have the common essential skeleton. The structure of "A" has the features that are two T-junctions, one cycle, and one acute curve. If our method effectively extracts essential skeleton graphs from various "A" images, the skeleton graphs must have these features. Figure 5A shows the skeleton graphs extracted from not-distorted "A", rotated one, and rotated and distorted one. These skeleton graphs had the common essential features.

However, skeleton graphs generated from handwritten character images or more distorted character images that include a same character may not always have same features. For example, the skeleton graphs generated from the handwritten digits "2" shown in fig. 5B was different from one shown in fig. 5C in spite of the same digit. The skeleton graphs shown in fig. 5B was the typical skeleton of "2". The skeleton graph generated from a printed character image will also have the same structure. The typical structure of the skeleton graph of "2" is one T-junction. While the skeleton graph shown in fig. 5C was not typical because the graph did not have the typical feature that was one T-junction. The skeleton graph of fig. 5C had two T-junctions and one cycle. These results suggest that the skeleton graphs generated from images including a same character may have different features.

4. Conclusion and future works

In this paper, we proposed a method to generate a skeleton graph representing essential features of a character in an image. We generated a skeleton graph from a character image using the GNG method, and then we deleted redundant edges of the skeleton graph using the RNG algorithm. The proposed method has been tested on images including a printed character, a distorted printed-character, a handwritten digit, and a noised character. The experimental results show the effectiveness of the proposed method. The skeleton graph preserved an approximation of the original shapes and had essential features of a character. The topology of the skeleton graph made by our method did not depend on rotation and distortion of a printed character.

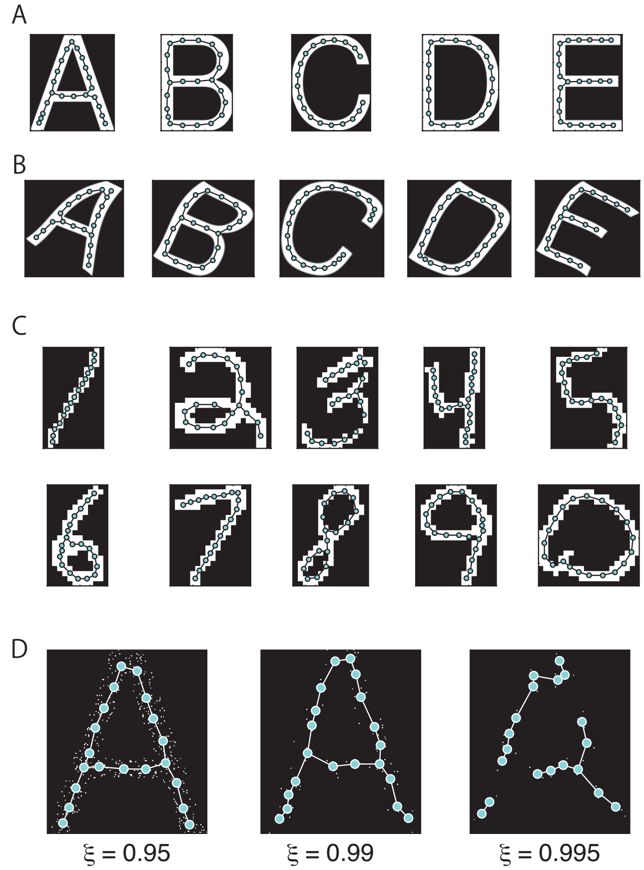


Fig. 4 (A) Alphabets. (B) Distorted and Rotated alphabets. (C) Handwritten digits. (D) Noised alphabets.

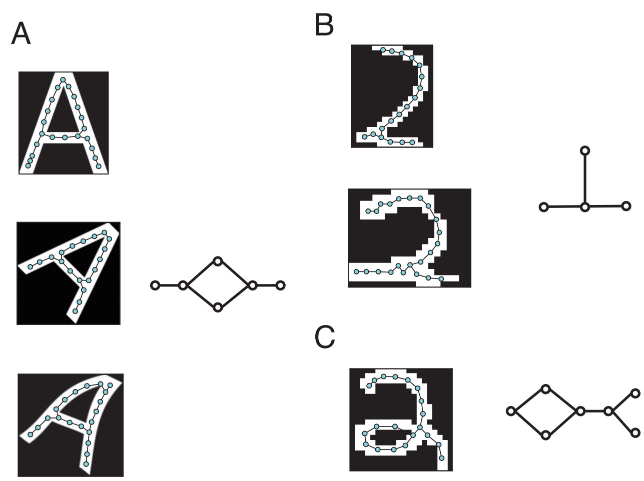


Fig.5 Relation between skeleton graphs and topology of the graphs. In the cases of A and B, the skeletons generated from images of same characters had same topology. In the case of C, the skeleton had different topology from one in B.

However, skeleton graphs generated from images including same character did not always have same features, for example handwritten characters. Furthermore, the skeleton graphs generated from images including different characters may have same topology. For example, the skeleton graphs made from "e" and "p" have one cycle and one T-junction, and the topology of these graphs is same. In this case, character recognition is not achieved using only topology of the skeleton graphs. To achieve character recognition, location of nodes, the number of nodes will be required.

In this study, we made a skeleton graph from a binary image. However, our method can be applied to directly making a skeleton graph from a gray scale image itself. To extract a skeleton graph from a gray scale image itself, the probability of selection of pixels on a character depends on the intensity of pixels in the GNG process.

In future work, we shall develop a character classification method using similarity of skeleton graphs because the skeleton graphs extracted from images including same character have similar features.

References

[1] A. Datta, S. K. Parui, and B. B. Chaudhuri, "Skeletonization by a topology-adaptive self-organizing neural network", Pattern Recognition, 34, 2001, pp. 617-629.
[2] E. Fatemizadeh, C. Lucas, and H. Soltanian-Zadeh, "Automatic landmark extraction from image data using modified growing neural gas network", IEEE Transactions on Information Technology in Biomedicine, 7, 2003, pp. 77-85.
[3] H. Frezza-Buet, "Following non-stationary distributions by controlling the vector quantization accuracy of a growing neural gas network", Neurocomputing, 71, 2008, pp. 1191-1202.
[4] B. Fritzke, "A growing neural gas network learns topologies", Advances in Neural Information Processing Systems, 7, 1995, pp. 625-632.
[5] K. Fukushima, "Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position", Biological cybernetics, 36, 1980, pp. 193-202
[6] J. García-RodríGuez, A. Angelopoulou, J. M. García-Chamizo, A. Psarrou, S. O. Escolano, and V. M. Giménez, "Autonomous growing neural gas for applications with time constraint: Optimal parameter estimation", Neural Networks, 32, 2012, pp. 196-208.
[7] Y. Holdstein and A. Fischer, "Three-dimensional surface reconstruction using meshing growing neural gas (mgng)", The Visual Computer, 24, 2008, pp. 295-302.
[8] L. Jia and Z. Hou, "An improved cost-efficient thinning algorithm for digital image", International Journal of Computer Science Issues, 10, 2013, pp. 190-197.
[9] T. Kohonen, "The self-organizing map", Proceedings of the IEEE, 78, 1990, pp. 1464-1480.
[10] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition", Proceedings of the IEEE, 86, 1998, pp. 2278-2324.
[11] J. Rocha and T. Pavlidis, "Character recognition without segmentation", IEEE transaction on Pattern Analysis and Machine Intelligence, 17, 1995, pp. 903-909
[12] G. T. Toussaint, "The relative neighbourhood graph of a finite planar set", Pattern Recognition, 12, 1980, pp. 261-268.

Emulating Trust Zone in Android Emulator with Secure Channeling

Arun Muthu[1], Rahim Rahmani[2] and Dinakaran Rajaram[3]

[1] **Department of Information and Communication Technology,
Kungl Tekniska Högskolan (KTH), Stockholm, Sweden**

[2] **Associate Professor, Department of Computer & Systems Sciences (DSV),
University of Stockholm, Stockholm, Sweden**

[3] **Department of Information and Communication Technology,
Kungl Tekniska Högskolan (KTH), Stockholm, Sweden**

Abstract

There is a raise in penetration of smart phone while using enterprise application, as most of them are downloaded from the public market, resulting in challenge for security framework, causing a threat to lose sensitive user data. To prevent this ARM introduces the virtualization technique in hardware level, which prevents processing of trusted application that is completely isolated from general processing. To improvise this, we need to understand ARM Architecture; however it is still black box for users and developers. In this article, we take a deep look at the hardware architecture of the ARM trust zone to study and analyze its implementation and also to create its replica in emulator. Moreover we describe feasibility of various designs, implementation of trust zone feature in android emulator; with sample trusted application called secure channeling and concludes with annotation of suitable design on future enhancement. The security domain for secure processing and utility in emulator is to benefit the user and developer community.

Keywords: *Trust zone, Emulator, Android, Virtualization, Security and secure channeling.*

1. Introduction

Technology seeking is expanding widely in all corners of the world and Smartphone is one among them. Now a days, we see plenty of smart phone users, progressively increasing in recent years and is expected to be more, since using smart phone is easy to access the application download, compatible and portable compared with laptop and notebook. However inability design and improper handling of the security critical functionalities of the Smartphone shows that no technology is resistant toward the security leak or attack till now. First android smart phone was introduced in the year 2008. Soon after the release, we found lot of security leaks and vulnerability in security architecture of the android OS. Since, Android Security Architecture, grants permission to perform any type of operation and So Google proclaimed that, “We tried really hard to secure Android. This is definitely a big bug. The reason why we consider it a large security issue is because root access on the device breaks our application sandbox."[1].

To solve this issue, service provider or smart phone marker adopts the new technology to solve this problem called trust zone [1]. Trust zone is the technology has gained wide acceptance and development in recent times. ARM trust zone is a hardware based system virtualization, help in handling third party application and security features in operating system. It consists of two zones called “Normal World” and “Secure World” [2]. The application which requires secure process will enter into secure zone from normal zone. There will be supervisor in the secure zone who will access the data from the normal zone and process it in secured way. The main functionality of the chip (ARM trust zone) will handle memory management unit, input and output guidance of the data, handling cryptographic keys and certificates etc. Since all these functionalities are internally organized, users and developer not aware of it. Android emulator is helpful for designing the business processes functionalities like application, transaction and payment etc whereas upcoming applications (trusted) for secure process cannot be resolved by the android emulator since, any support for software virtualization in it.

In this paper, we propose a design framework for emulated trust zone for android emulator. Our emulated trust zone was designed based on the important attributes used in ARM trust zone design. Furthermore, our designed system also handles the attributes in same way like ARM trust zone. So, it works as a replica for hardware trust zone chip in emulator.

The main contributions of our work can be summarized as:

- Reviewing current practices and theories on implementation of ARM Trust zone.
- Analysis and design the appropriate model by comparing the actual working of trust zone in hardware level with design and idea of emulated one.
- Create a trusted application of our choice matches with design criteria.
- Proposing step-by-step approaches to solve the research tribulations.

2. Problem Statement

In this section, we describe the problem and motivate the need for information to design the emulated trust zone. We are primarily concerned on the emulation of trust zone in virtual emulated world. The main drawback is we cannot focus on single model or design. So, we have to compare different entities like real smart phone trust zone with android emulator functionalities. The practical difficulties in knowing the attributes of ARM trust zone and design & develop the software module matching to it.

The list of attributes used in ARM trust zone [2] need to be replicated:

- Secure memory management
- Monitor mode and supervisor mode
- Interrupts
- User space
- Trusted Application with secure channeling

Table 1. Difference between ARM trust zone and emulated trust zone

	ARM trust zone	Emulated trust zone
Secure Memory Management	Created during booting time	Created during execution after SCM call
Monitor mode	NS and S bit value changes according to transit between the worlds	NS and S bit value changes according to transit between the worlds
Interrupts	Operating system takes care of it	Operating system takes care of it
User space	Application start from user space and it enter into kernel space	Same as ARM trust zone functionalities
Trusted Application	Loaded during boot time and it's static	Loaded during execution time and its dynamic
SCM Call	ARM instructions [8]	Procedure call
Debugging	We cannot debug application since it already compiled	Here, Native C code in JNI layer user" *.so" file
Registers	ARM registers	Variables

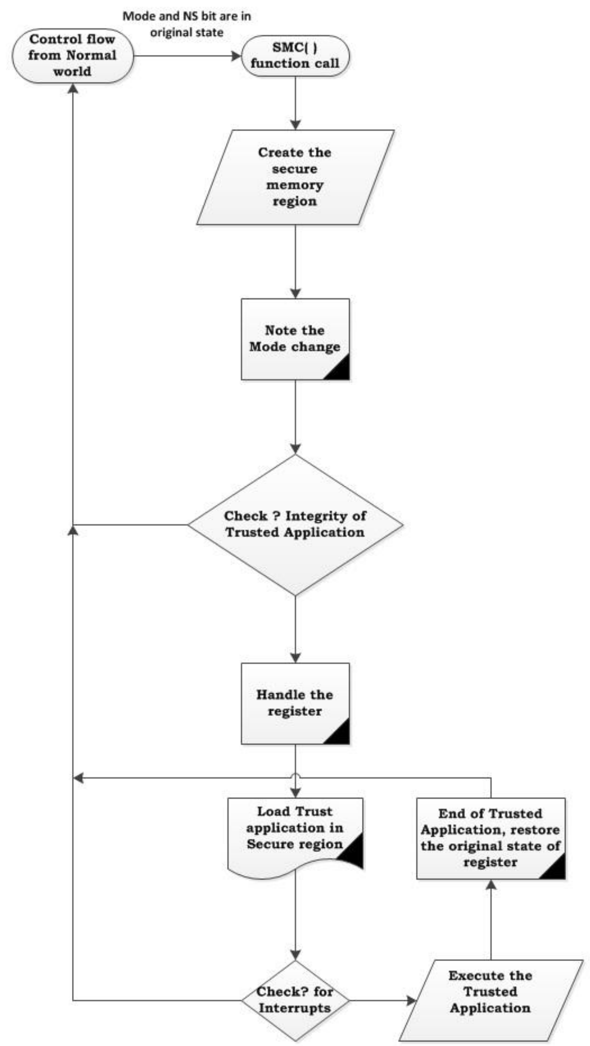


Fig. 1. Assumed control flow

By considering all these attributes, we can implement emulation in two ways.

- Top level emulation (above driver)
- Low level emulation (below driver)

2.1 Explanations about these two types of emulation

- We continue with the SMC (Secure monitor call) emulation track which gives us a clean/empty secure world that we have to fill with executable code and help to create our own Trusted environment (e.g. Global Platform TEE)[2].
- We create our own TZ driver level, to load trusted application in secure world after SMC call and exchange between the worlds [7].When we try to emulate SMC instruction – we need to replace it with the suitable function (system call) which is targeting the particular address location of the trust application (normal C function call) in the file system.

Problems in Low level emulation:

- Low level emulation is not achievable, due to the compatibility problem between the ARM board (phone) and Intel board(host machine).In real phone there are two different memory unit and accessing point to prevent and guide the execution flow, where as in emulator, single memory strip and single mode operation.
- The trusted code is split into two parts, when loaded by the boot ROM. The first part is regarded as trusted application persistent in secure memory location during the control flow. The other part is initialized or triggered from user space (application). This indirect leads to undefined problem, that is trusted code cannot be unloaded or reloaded [8].
- Trusted application (TA) uses the functionality of the Secure ROM API. Secure ROM API is the main API to be used by trust application (TA) in ARM. So, it is not possible for dynamic trusted application to use defined functionality. So TA should always be static and specifications are loaded in boot time itself.
- Accessing global variable is difficult. No firmware is available to execute or support it.

Problems in Top level emulation:

- Main problem is to create our own execution flow– Android application, Trust zone Driver (TZ Driver), SCM mocking code and trusted application.
- We need to create the fake monitor code i.e. initialization of trusted application and its properties (key, ID, Data, flag and address etc) is mocking the boot ROM specification.
- Handling register functionalities are difficult, since no CP15 register [Appendix] for secure and non secure banking. So, we need to assume and design our own apk layers and TA to perform secure world operation.
- Some verification of trusted application need to done before calling driver codes. This indicates that TA is loaded dynamically and led to problem of in granting permissions to the user.
- Coping TA to specify address (secure memory region) has no meaning in it. Since, we need to bring realistic view of ARM processor.
- No details about secure ROM API: so we create monitor mode as variable or flag status to indicate the user or programmer that, CPU mode has been changed and control flow is switched to secure processing. By reviewing all these entire problems, we prefer, Top level approach is more relevant in this case.

3. Various Designs

By considering these problems with both levels, we suggest top level matches which have feature specifications for replicating the ARM trust zone [5]. Those are:-

- Supervisor design
- Dual memory design
- Static memory design

3.1 Supervisor Design

In this design, we use two instance of kernel layer (i.e.) normal and secure. Supervisor is responsible for the analysing the instruction and then send them into appropriate world. This shows that, supervisor is responsible for context switching between the worlds. The supervisor design is more efficient since it differentiates the two kernels layer and therefore the two worlds. However implementing this solution is more difficult, since one software module will need to understand the already existing and compiled code [5].

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 1, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

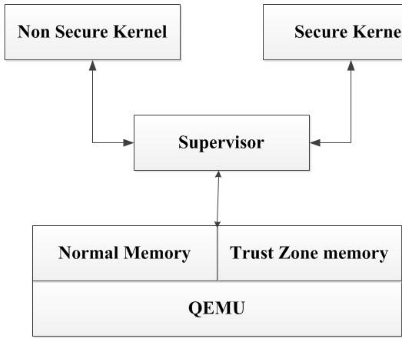


Fig. 2. Supervisor design

3.2 Dual Memory Design

In this design, we create two memory management units to manage each kernel layers to support previous design (supervisor design). As like previous design, here two memory units represent two worlds. The secure memory unit is responsible for the context switching between the worlds. The normal memory unit, will access the part of secure memory unit in order to convey the results between them. So, this indirectly leads to creation of the shared memory region.

This design is not so efficient to isolate the two worlds, even though it solves the supervisor design problem i.e. creation of software module to the stack. Handling memory unit makes this implementation of design more complex [5].

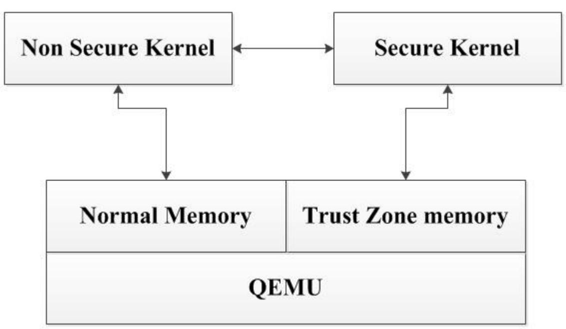


Fig. 3. Dual memory design

3.3 Static Memory Design

By seeing all these complexity in the nature of design, we will stick to one kernel and memory unit. Isolation of secure world can be achieved by creating the static memory in the section of physical memory. The Kernel will monitor each instruction and will forward it according to the appropriate memory region. The kernel is responsible for the context switching. The secure world can access the data from both world and communicate results to normal world without any shared memory concepts [5].

This design is much easier comparatively to other design mentioned above. However, the main problem is handling security features (i.e.) making static memory region has secured one.

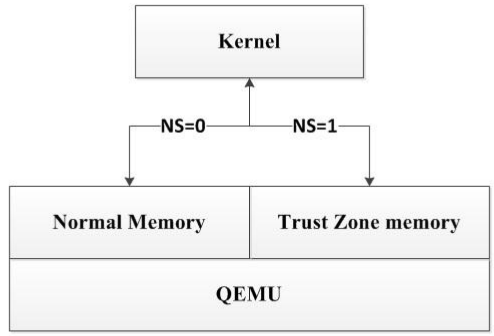


Fig. 4. Static memory design

4. Design Framework

In this section we describe the feasibility abstractions that enable the high level design specification of trust zone in emulator. By clearly viewing the design specification problem, we prefer top level approach with static memory design is more relevant in designing the emulator trust zone in emulator.

Benefits of top level design:

- working as standalone program
- debugging is quite easy
- Navigation from APK->Driver->SCM_CALL->TA is notable.
- We can protect the memory region (secure memory) using mmap methods.

```
/* buffer is temporary variable to
store executable codes */
buffer=mmap(NULL,cmd.Size,PROT_EXEC|PRO
T_READ|PROT_WRITE,
MAP_PRIVATE|MAP_ANONYMOUS,1,0);
```

4.1 Design impact of Top level design on attributes of the emulated trust zone:

Secure memory management.

We followed the same principle and design to develop the trust zone features in an emulator. But there is slight modification in the design, when we consider the same approach in emulator. There is no special hardware support for the work flow. Since, in real target there is concept called shared memory between the processor and memory units, but it is missing in emulator. So, we redesign the resource utilization as per emulator accordingly with same NS bit.

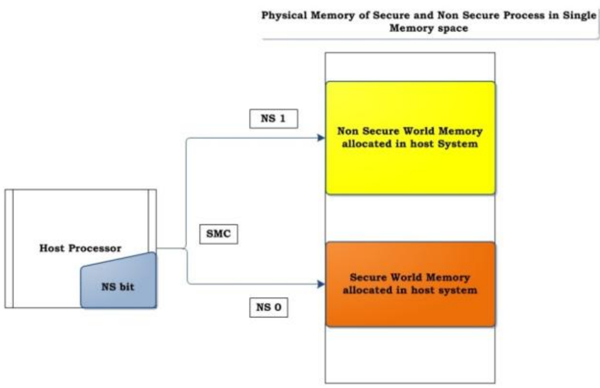


Fig. 5. Secure memory management design for emulated trust zone

Monitor Mode and supervisor mode.

In case of emulator, SMC can be replaced by the function or procedure call.SMC is ARM instruction, which will not work in emulator. SMC function call helps to prevent the non secure state in accessing region of physical memory, since each states operates on own memory address space. Mode changes are captured by the variable values in the programming logic and tracking them are also be done.

Table 2. Bit Value

	Monitor Mode	Supervisor Mode	Non Secure bit	Secure bit
Application in Non secure region	0	1	1	0
Application in Secure region	1	1	0	1

The general purpose registers and processor status register are not blanked between the secure and the non secure states. When execution switches between the non secure and secure states, ARM expects that the values of these registers are switched by a kernel running mostly in monitor mode. Whereas, system coprocessor register are banked between the secure and non secure security states. A banked copy of a register applies only to execution in the appropriate security state.

Interrupts.

Many uses of the security extension can be simplified if the system is designed so that exceptions cannot be taken in monitor mode. Setting bits in the secure configuration register causes one or more of external aborts, IRQs and FIQs to be handled in monitor mode.

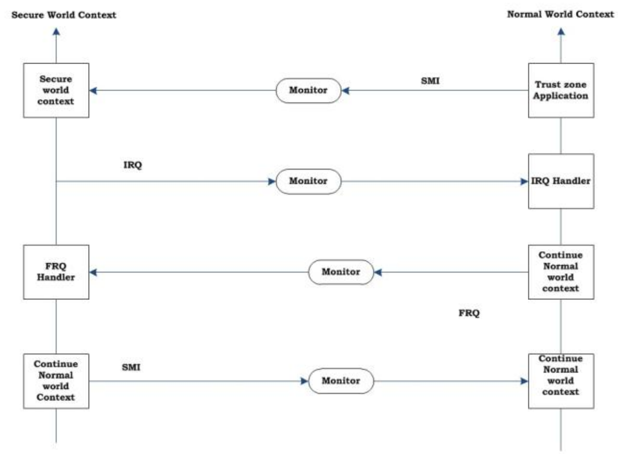


Fig. 6. Interrupt handling

If an exception is taken in monitor mode of non secure state, the Secure Configuration Register (SCR) [Appendix] bit is set to zero [8]. This indicates the operating system that exception occurred. However, if an exception is taken in monitor mode in secure state then, register bit is not set to zero.

User space.

User space is the main feature of the top level design, since the control flow start from it. It helps to delegate operation to an authorized domain. The applications have specific system call to the operating system kernel. This include syscall numbers, interfaces etc. In general, environment for secure world user space application should be simple system call interface to support C run time libraries and compiler tool chains.

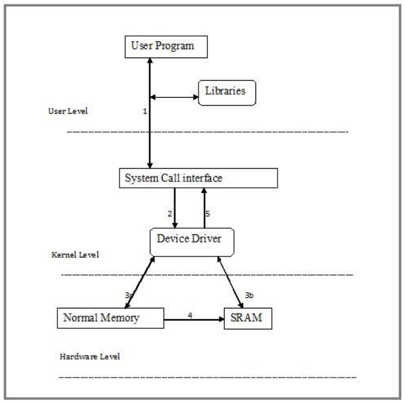


Fig. 7. Control flow from user space

1. Application enter into the user level (android *.apk to system interface)
2. System call leads to corresponding driver file
3. These are things are connected by the JNI- Java native interface.
    3a. initially all process start in normal memory
    3b. whenever the special treatment is required during the execution, control will be directed to secure memory region via driver file when SMC procedure call is made from application.
4. Transfer the block of data from normal memory to secure memory which require the special secure treatment/Process.
5. Once, a process is over, session is closed and returns to normal memory space to further execution.

This process is repeated until process requires special treatment/secure computing

Trusted application.

According to our design, trusted application is the place, where actual business case is introduced to solve and end of the control flow from Top level (apk). For this paper, we constructed the application to establish the secure channel [3] [6] communication between host and external world (for: server of the business provider). So, by replacing the code with desired business logic we can achieve the corresponding significant output.

Why secure channeling?.

In real world application, trust zone works with support of both internal architecture as well as external world. Consider an example of banking transaction, mutual authentication is required by both user and bank server in order to establish the communication between them. This is can be achieved by the "challenge response" mechanism.

Passwords can be reused which may lead to compromise the entire communication. So challenge response is transmitting passwords change each time. Encrypting those passwords with key make the communication more secure.

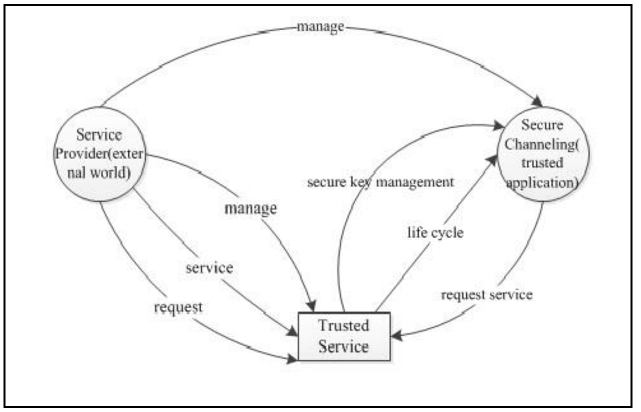


Fig. 8. Trusted services

In smart phone, hardware supports comes with, secure memory region to execute the mechanism (secure channeling) and application created by external service provider. So the user treats black box since he unaware about the mechanism of application. Some of the mechanisms controlled by phone manufacturers are

- Resource of hardware architecture and environment for software execution
- Installing additional application requires permits and assistance
- Billing and usage management are controlled by network operators
- Service management are controlled by service providers so subscribers are not in position to select or change the service

In our case, we use the same principles to design our trusted application. But due to these entire problems mentioned above, we stick to the basic authentication with standard encryption methods to establish the communication between the host and external world to supports our design matches with real world mechanism.

We assume in trusted application mechanism of establishing the secure channel between the mobile and external world is handled. So, trust zone handle the mechanism and input data. Secure channel means mutual authentication between the host (android emulator) and external world. This can be achieved by two way steps.

- Creation of challenge response
- Comparing cryptograms

Initially, we need to create the mutual authentication mechanism, which can be understood by the host and external world. This mechanism is called as challenge response [4].

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 1, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

Random block of 8 bytes should be created by both the host and the external world application called as the host and external challenge. Derivation challenge data of 16 bytes should be formed by the combining host and external challenge [3] [7].

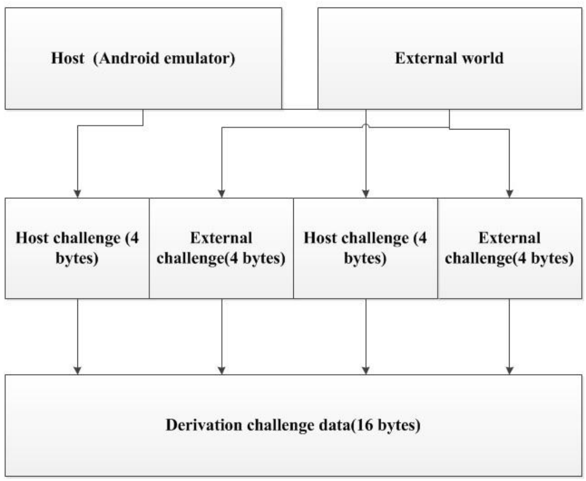


Fig. 9. Creation of derivation challenge data [3] [7]

Encrypt the derivation challenge data with Static encryption key to form session key. By using the session key, cryptograms are produced. Input data (credentials, pin numbers etc) are combined with the host challenge and send to external world application.

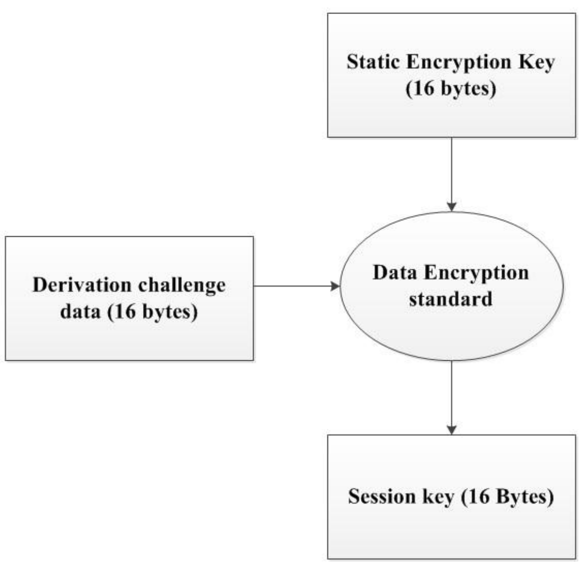


Fig. 10. Session key creation

External world application checks host cryptogram and compares it with his own cryptogram (external) generated by the same operation with data from host cryptogram. Secure channel is established if the both the operation leads the same result. So our assumed of communication between host and external world with secure channeling after SCM call will be:

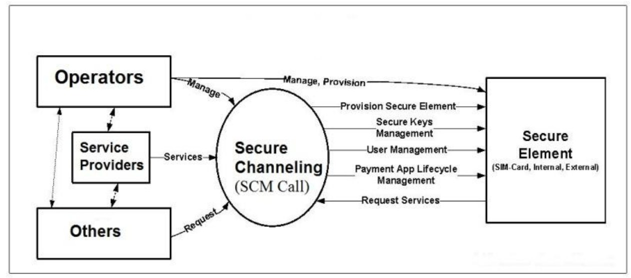


Fig. 11. Communication between host and external world [6]

5. Validating System Model

In this section, we describe our experiments to evaluate and validate design framework. In order to achieve the design of emulated trust zone, we have shown the evidence in analyzing various designs. As mentioned in earlier chapters, Model consists of two layers.

Apk creation or top layer design, defines only certain set of android function call which call the native c procedure which connects the kernel to top layer.

Layer 1: SCM function, actual communicating elements that include driver and procedure for connecting layer 2

- Secure memory creation
- TA creation

Layer 2: Secure channeling, defines the communication between the TA and external application.

5.1 State transition

Considering our example, there according to labelled transition system are 3 stages s0, s1 and s2, where s0 is normal state, s1 is secure state and s2 is external state communication. Process implicitly by calling initial state s0 and follow its successor state S=2T. So the successive states can be defined as successive state(s) = {(s, a, s1)|(s, a, s2) $\in$ T} where T is transition of states. Since SCM call is the initial state of the process and simple "C" procedure for implementation make it straightforward verification algorithms.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 1, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

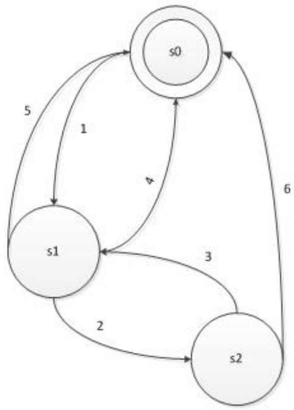


Fig. 12. Transition states

1 & 4: S0 to s1 – transition from top layer APK to SCM call or inturns normal world to secure world.
2 & 3:S1 to s2 – secure channelling communication between host (TA) and external world
5: when interrupt occur in secure processing, control will be revert back to normal world.
6: when interrupt occur in communication between host and external world also, lead back to normal world.

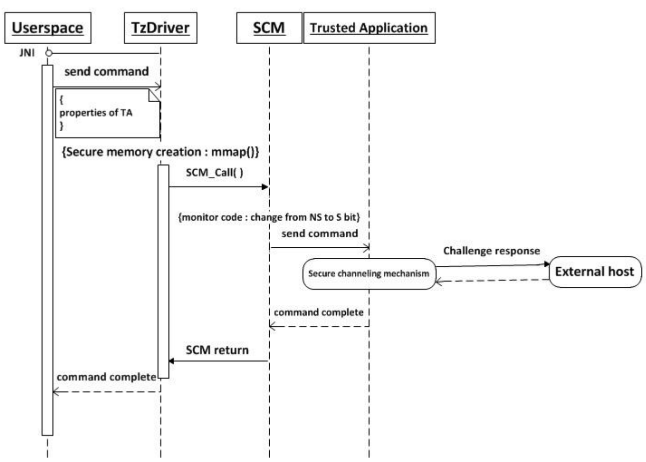


Fig. 13. Control flow from user space to SMC call

5.2 Order of algorithm

Order of the algorithm we used in SCM call is very simple and linear functionalities are used. So the order of the program will O(n2).

Pseudo code with order O(n2):.

```
//CPU_common.h
// we need to define the parameter used by the SRAM
Typedef unsigned long sram_addr_t; // this way we can create data type of SRAM
Sram_addr_t size, offset; // this parameter will be used by SRAM, to describe its characteristic
/* In same file, we need to create the definition for the methods/Functions used by the QEMU, like allocating the memory, releasing the memory of SRAM */
Void qemu_sram_alloc (sram_add_t size);
Void qemu_sram_free (sram_addr_t size);
// implementation of this methods can be viewed in android_arm.c file

//Android_arm.c
Void qemu_sram_alloc (sram_add_t size)
{
/*This will allocate the memory (say for example 512 Mega bytes) other than memory create by emulator for normal function*/
}
Void qemu_sram_free (sram_addr_t size)
{
/* whenever, application is shut down/ Close memory used by the SRAM should be de allocated */
}

// tee_sram_driver.c
/*this file will be core of the SRAM feature. This only decide to access the data in the particular memory location */
/* Methods used for the Tee specification to handle the SRAM function like read, write, open, close etc */
Tee_sram_open (); // open the driver
Tee_sram_close (); // close the driver
Tee_sram_read (); // for read operation
Tee_sram_write (); //for close operation
Tee_sram_session (); // for creating the session
Tee_sram_reset_session (); // reset the session
Memrefs_normal_to_sram (); // transfer of data from normal memory to SRAM memory space
Memrefs_sram_to_normal (); // transfer of data from SRAM memory to normal memory space
```

But in case of secure channeling algorithm, order of the growth will be larger since we are using hashing, crypto functionalities. So, the order will be Ω(n2)

Pseudo code with order Ω(n2):.

```
public boolean establishSecureChannel
() throws Exception {
//host
    byte [] hostChanllenge = new byte[]
{};
//card
    byte[] cardChanllenge = new byte[]
{};
};

//session key creation
sessionKey =
deriveSessionKey1(hostChanllenge,
cardChanllenge, KDC_enc_data);

//Cipher the text
Cipher cipher =
Cipher.getInstance("DES/ECB/NoPadding")
;
```

5.3 Static analysis

Static analysis is very efficient way for finding some interaction (in static memory design), may lead to non determinism state .In simple terms called as overlap finding. It depends on structure of the design. By manual examine of the code functions, overlapping makes the statement divergent.

Table 3. Static analyses

	APK	Driver	SCM	Secure channeling
APK	--	X	X	X
Driver	X	--		X
SCM	X		--	
Secure channeling	X	X		--

5.4 Dynamic analysis.

Dynamic analysis is process of proving functionalities (Static memory design) of application in more logical way. Consider, there are two trusted application A1 and A2 with the functionalities f1 and f2. So, that A1 acquires f1 and A2 have f2.By consider one of the test cases of this project: “Trust zone can access only one process/application at a time”.

So, A1f1 $=\emptyset$ where as A1f1A2f2 $\neq \emptyset$, where $\emptyset$ is the property of trust zone.
It is easy to understand all this features interact with property. We are considering only properties associated with functionalities like f1 and f2 related to $\emptyset$1 and $\emptyset$2. So we need more selective approach without violating the proprieties $\emptyset$i for (f1…fi)Ai.

5.5 Feasibility study.

By considering all these parameters and factors mentioned above, with all the three designs of analysis, we formulate the table to with pros and cons of each design.

Table 4. Evaluation of design based on Attributies

Factor	Supervisor design	Dual memory design	Static memory design
Secure memory management	No separate memory handling for accessing two different kernel layer	Two isolated memory area to handle to two kernel region	Single strip of memory which is divided into normal and secure memory region
Monitor mode	Supervisor will handle the control flow between two kernel region	It work same as Supervisor design, only different is supervisor need to handle control flow between two memory	Supervisor is just a bit variable.
Trusted application	It accessed first in normal kernel and then moved to secure kernel	Here, Application is accessed by normal memory first and then moved to secure memory	Trusted application is copied from normal memory to secure memory by changing the bit value of NS and S bit
SCM_CALL	Supervisor will call for secure processing	Its work same like supervisor design	It just the procedure call to inform application require secure processing
Interrupts	Handled by Operating system	Handled by Operating system	Handled by Operating system
Debugging	Since all process happen in kernel space. No chance of debugging	Since all process happen in kernel space. No chance of debugging	By using NDK, we can debug the user space application but not in kernel space
Register	Internal registers	Internal registers	Variables

Table 5. Evaluation of design based on Functionalities

Factor	Supervisor design	Dual memory design	Static memory design
Complexity	High : Since switching between kernel is not easy	High : memory handling and switching is difficult	Low: since all process happen on top level(user space) switching and handling memory are easy
Reuse	No possibilities of re use of code	No possibilities of re use of code	User space application can be re used. We need to change the kernel driver according to user space program
Implementation	Implementation is not easy since two kernel layer involved	Implementation is not easy since two memory layer involved	Implementation is easy
Security	Security features are difficult to handle here	Protecting the memory is highly difficult	We can use any security standard to encrypt/decrypt the communication between host and external world

By seeing above all, Static memory design would more feasible compare to other two. From various evaluation and implementation criteria, we found out the actual working of ARM Trust zone as well as various ways of implementing replica of trust zone (static memory design) in emulator.

We designed the emulated trust zone in such a way to give justification to the real working mode, even trusted application and secure channeling are also designed to provide support to the design. Finally, we have consolidated artefact in descending order, so that reader can easily follow the flow and idea behind each chapter and design of it.

6 Conclusions

Our aim is to develop a software implementation of trust zone in emulator. To achieve this we have analysed the existing design, understanding the feasibility of implementation of those design in emulator. After exploring alternate design of trust zone, we come to the conclusion that, this research finds the solution to make an architecture design to implement trust zone in android emulator with merely justification to real ARM trust zone design. As the result, various architectural designs of trust zone are discussed and best one is chosen. So, the main goal for paper is drawn.

- Handling Memory segment
- Driver files
- Swapping programs to the specific memory
- Handling system call and CP15 [Appendix] variables
- Sample trusted application – Secure channeling

On overall design, our framework starts with android application, navigate to JNI an interface, to provide high level platform independent experience to the developers. Based on assumption, we implement the small example called “secure channeling” that provide the construction and emulation of trusted application running in secure memory region. SCM call is made in kernel layer to communicate with external host (e.g.: server–client program) to provide mutual authentication mechanism.

- Alternate designs and control flow of framework will explain the overall feasibility for the construction of trust zone in android emulator. In our case, we justify that static memory design is more feasible comparatively with other designs.
- Since, we choose static memory design for our case, so separate memory region is created for secure channeling and context switching is the navigation of control from normal memory to secure memory region and vise versa. Meanwhile other features like interrupts, application handling also done to support context switching.
- Here secure channeling [3] is the process of authenticating the host and external world by using Challenge response. This can be extended by using cryptography standards in order to provide more justification to the design.

Future enhancement & significance.

- By bringing Trust zone (static memory design) into software emulation, have plenty of significance for secure processing application and great contribution to Google android development.
- Environment can be extended to all applications required secure treatment to process and testing

of hacking methods over the applications processed in trust zone.

- This project also helps with application of Trust zone emulation. So, it will avoid hardware limitation on the security of the application for open source world.
- Implementing personalized and complied trusted application – FOTA, Banking application etc.,
- Implementing real ARM SCM call in emulator
- Designing more alternate design for trust zone
- Dependencies on hardware will be reduced
- Help future researcher to work on trust zone and security feature of it.
- Successor of Wallet project of Google

Appendix

Table 6.CP15 Register

Register	Description
C0	Main ID Register (MIDR)
	Cache Type Register (CTR)
	TCM Type Register(TCMTR)
	TLB Type register (TLBTR)
	Multiprocessor Affinity Register (MPIDR)
	Processor Feature Register
	Debug Feature Register (ID_DFR0)
	Auxiliary Feature Register(ID_AFR0)
	Memory Model Feature Register
	Instruction Set Attribute Register
	Cache Size ID Register(CCSIDR)
	Cache Level ID Register(CLIDR)
	Implementation Defined Auxiliary ID register(AIDR)
	Cache Size Selection Register(CSSELR)
C1	System Control Register (SCTLR)
	Implementation Defined Auxiliary Control Register(ACTLR)
	Coprocessor Access Control Register (CPACR)
	Secure Configuration Register (SCR)
	Secure Debug Enable Register(SDER)
	Non Secure Access Control Register(NSACR)
C2	Translation Table Base Register 0(TTBR0)
	Translation Table Base Register 1(TTBR1)
	Translation Table Base Control Register (TTBCR)
C3	Domain Access Control Register(DACR)
C4	Not used
C5	Data Fault Status Register(DFSR)
	Instruction Fault Status Register(IFSR)
	Auxiliary Data and Instruction Fault Status Registers (ADFSR and AIFSR)
C6	Data Fault Address Register(DFAR)
	Instruction Fault Address Register(IFAR)
C7	Cache and Branch Predictor Maintenance Functions
	Virtual Address to Physical Address Translation Operations
	Data and Instruction Barrier Operation
	No Operation (NOP)
C8	TLB Maintenance Operation
C9	cache and TCM Lockdown Register and Performance Monitor
C10	Memory Mapping and TLB Control Register
	Primary Region Remap Register(PRRR)
	Normal Memory Remap Register(NMRR)
C11	Reserved for TCM DMA Register
C12	Security Extension Register
	Vector Base Address Register(VBAR)
	Monitor Vector Base Register(MVBAR)
	Interrupt Status Register(ISR)
C13	Process Context and Thread ID Register
	FCSE Process ID Register(FCSEIDR)
	Context ID Register(CONTEXTIDR)
	Software Thread ID Register
C14	Not used
C15	Implementation Defined Register

Acknowledgments

The alacrity shown by Ms.**Anjani Dubey**, Google Hyderabad & Mr. **Karthikeyan S**, NAL India, has gone long way in improving the form and content of this article.

References

[1] David, Kleidermacher. *Bringing Security to Android based System*. [Online][Cited: June 10, 2013.] http://www.iqmagazineonline.com/current/pdf/Pg56-58_IQ_32-Bringing_Security_to_Android-based_Devices.pdf.

[2] **ARM Security Technology.** *Building a Secure System using TrustZone Technology*. [Online] [Cited: June 13, 2013.] http://infocenter.arm.com/help/topic/com.arm.doc.prd29-genc-009492c/PRD29-GENC-009492C_trustzone_security_whitepaper.pdf.

[3] **Rajaram, Dinakaran.** *Secure Over the Air (OTA) Management of Mobile Applications.* ICT, KTH Royal insitute of Technology. Stockholm : s.n., 2012.

[4] **C. Newman,A. Menon-Sen, A. Melnikov & N. Williams.** *Salted Challenge Response Authentication Mechanism* (SCRAM). [Online] 2010. [Cited: June 17, 2013.] http://tools.ietf.org/html/rfc5802#page-20.

[5] **Alexander M.. Frisvold, Alex Meyer, Nazmus Sakib, Eric Van Buren.** *Android Security. Iowa State University: s.n.,* 2012.

[6] **Zhao, Hao.** *INTEGRATED SECURITY PLATFORM FOR MOBILE APPLICATIONS.* KTH School of Information and Communication Technology, KTH Royal institute of Technology. Stockholm : s.n., 2011.

[7] Global Platform. [Online] [Cited: June 21, 2013.] http://www.globalplatform.org/showpage.asp?code=specifications.

[8] Winter, Johannes. *Trusted Computing Building Blocks for Embedded Linux-based ARM TrustZone Platforms.* Institute for Applied Information Processing and Communications (IAIK), Graz, University of Technology. Graz : s.n.

About Authors

Arun Muthu is an Information Security Manager in Feeders Technologies, India founded in 2008. He is responsible for providing security solution to the corporate business and research. He works closely with feeders leaders in driving the growth & expansion of the company's revenue & market. He holds Master degree in Information & Communication System Security from Kungl Tekniska Högskolan (KTH), Sweden 2012. His master degree and his earlier background at Cognizant and Sony Mobile provide him with a solid foundation for research innovation in recent years. His areas of interest are Cryptography, Secure Software Engineering, Privacy, and Trusted Computing.

Dr. Rahim Rahmani is an Associate Professor of Computer Science and system science in University of Stockholm from 2012, focusing on QoS and optimization for management of cloud computing infrastructure and wireless sensor networks, such as OpenFlow. He earned a technical doctorate in adaptive Active Queue Management (AQM) algorithms for access routers in heterogeneous networks. He has served as reviewer and in technical committees of international conferences. He is a member of the editorial board of International Journal of Wireless Networking and Communications.

Dinakaran Rajaram joined Feeders Technologies as Researcher in Information Security as well as IT Security Auditor from April 2013.He obtained his Bachelor of Technology (Information Technology) degree in 2009 from Thiagarajar College of Engineering, Madurai and Master of Science degree in Information & Communication System Security from Kungl Tekniska Högskolan (KTH) in 2012. He is recipient of "Erasmus Mundus" scholarship for his master studies. His researches interests are Security Protocol design, Trusted Computing, Identity Theft, Cryptography and Secure Channeling.

A speech recognition based approach for development in C++

Mubbashir Ayub[1] , Muhammad Asjad Saleem[2]

[1]Faculty of Telecommunication & Information Engineering
University of Engineering & Technology Taxila, Pakistan

[2]Faculty of Telecommunication & Information Engineering
University of Engineering & Technology Taxila, Pakistan

ABSTRACT
Software development using programming languages requires keyboard for input and all programming languages are mostly text oriented. This text oriented nature of programming languages is a barrier to persons suffering from arms disability. A person having brilliant mind and potential for programming skills, but suffering from arm injuries or being disabled could not become a programmer. To be a good developer a human must memorize the syntax and keywords of a programming language. In our research work we propose a methodology for C++ programming language where a programmer will speak and code in C++ will be written accordingly. Structure of special program constructs will also be created simultaneously.

Key words:
Automatic speech recognitions, Speech API, Shared Recognizer, Object Tokens, Text to Speech

1. INTRODUCTION

Programming environments can create frustrating barriers for the growing numbers of software developers that suffer from repetitive strain injuries (RSI) and related disabilities that make typing difficult or impossible. Not only is the software development process comprised of fairly text-intensive activities like program composition, editing and navigation, but the tools used for programming are also operated textually. This results in a work environment for Programmers in which long hours of RSI-exacerbating typing are unavoidable.

Grappling with the syntax of a programming language can be frustrating for programmers because it distracts from the abstract task of creating a correct program[1]. Visually impaired programmers have a difficult time with syntax because managing syntactic details and detecting syntactic errors are inherently visual tasks. As a result, a visually impaired programmer can spend a long time chasing down syntactic errors that a sighted programmer could have found instantly. Programmers facing repetitive stress injuries can have a difficult time entering and editing syntactically detailed programs from the keyboard. Novice programmers often struggle because they are forced to learn syntactic and general programming skills simultaneously. Even experienced programmers may be hampered by the need to learn the syntax of a new programming language.

The primary tool used for programming is a specialized text editor [2]. Early text editors were manipulated entirely via keyboard, and leveraged many keyboard shortcuts for common operations in order to speed the programming process. Unfortunately, keyboard shortcuts are usually composed of an alphanumeric key combined with one or more modifier keys (e.g. Control, Option, Alt), which contributes to RSI when keyed unergonomically. These days, text editors are usually embedded in integrated development environments (IDEs) that provide the programmer with integrated services, such as compilation and debugging [1].

One way to reduce the amount of typing while programming is to use speech recognition. Speech interfaces may help to reduce the troubles of RSI among computer programmers. At the other side speech programming may increase access for those already suffering motor impairments. Many disabled programmers are already bootstrapping voice recognition into existing programming environments [3]. However, speech does not map well onto the available applications and programming tasks. Our research uses a principled approach from field of programming languages to allow developers to use speech with much more powerful control.

We come up with a solution that addresses the above mentioned problems. Our solution requires a microphone attached to the computer. User should have all the basic understanding of C++ language. He will speak his code that will be in English language but conforms to C++ syntax and semantics. Our system will only provide code writing feature. After writing code user will copy and paste this code to any C++ compiler and will compile code in that compiler.

2. OVERVIEW

Our system is based on Microsoft platform. System is developed using Microsoft Visual studio 2010. Windows speech recognition is used to capture voice and Microsoft Speechlib is used to convert voice to text. This converted text is then used to generate C++ code. Section 3 describes the system architecture in detail. Section 4 states the system comparison with other existing systems. Section 5 concludes our work with discussion of future directions.

3. SYSTEM ARCHITECTURE

Our system is divided into the five modules as shown in [Figure.1].

- Graphical User Interface (GUI)
- Voice to Text Converter
- Code Generator

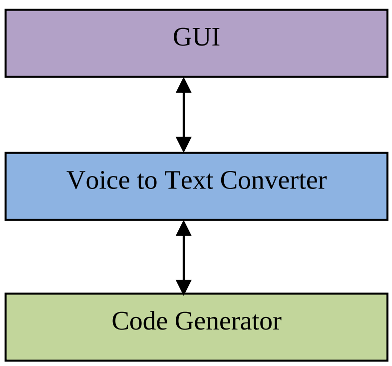


Figure.1 System Architecture

Before getting into the details of the system, let's have a brief overview of the high level working of our system. GUI module provides an interface for users to interact with the system. Voice to text converter is responsible for converting each listened word to text. A user will speak code using microphone and text converter will convert this spoken code to text. For this converted text C++ syntax and semantics are applied to this text to generate standardized C++ code.

3.1 Graphical User Interface (GUI)

GUI gives the visual appearance of the virtual file system to the end user. GUI color schemes, layout, working and behavior are quite similar to Windows Explorer. Windows XP style task pane provides easy access to common operations and gives appealing look. Standard Toolbars, popup menus and shortcut keys make operation of software easy for all type of users. Easy to Use, Easy accessibility to functions and Appealing appearance are the main features of GUI.

3.2 Voice to Text Converter

The core part of our system is voice to text conversion. We used Windows Speech Recognition(WSR), which is built on top of SAPI, for capturing voice and Microsoft Speechlib API for converting this voice to text.
SAPI version 5.4 is shipped with windows 7 and supports two distinct types of speech recognition; dictation and command and control. In our research we used dictation type of speech recognition. In this type of speech recognition machine listens to what we say and attempts to translate it into text. The accuracy of dictation ties directly to the CPU's speed and the system's available memory. The more resources, the more contexts that can be considered in a reasonable amount of time the more likely the resulting recognition will be accurate.
SAPI 5.4 supports two types of recognizers inprocess recognizer (SpInprocRecognizer) and shared process recognizer (SpSharedRecognizer). The inprocess recognizer claims resources for the application, so, for example, once an inprocess recognizer claims the system's microphone, no other application can use it. A shared recognizer runs in a separate process from the application and, as a result, it can be shared with other applications. This allows multiple applications to share system resources (like microphone). In our application we are using shared process recognizer because shared recognizer allows an application to play nicely with other speech enabled applications on system.
A recognition context is an object that manages the relationship between the recognition engine object (the recognizer) and the application. A single recognizer can be used by many contexts. For example, a speech enabled application with 3 forms will likely have a single engine instance with a separate context of each form. When one form gets the focus its context becomes active and the other two forms contexts are disabled. In this way, only the commands relevant to the one form are recognized by the engine. A single recognizer can be used by many contexts. For example, a speech enabled application with 3 forms will likely have a single engine instance with a separate context of each form. When one form gets the focus its context becomes active and the other two forms contexts are disabled. In this way, only the commands relevant to the one form are recognized by the engine. SAPI is smart enough to create the shared recognizer object for us automatically when the SpSharedRecoContext is created. In our scenario we are using dictation type of speech recognition. For this purpose we created a grammar object and load the grammar with SLOStatic value to set the dictation top of grammar as static. To set this grammar object to use dictation type of speech recognition we initialize SpeechRuleState state property of grammar object to SGDSActive.

In recognition event handler the ISpRecoResult interface is used by our application to retrieve information about the SR engine's hypotheses, recognitions, and false recognitions. The most common use of the ISpRecoResult interface is retrieval of text recognized by the Speech Recognizer. The ISpRecoResult interface also supports the retrieval of the original audio that the SR engine recognized. An application can set interest in the SR engine's failed recognitions by calling ISpEventSource::SetInterest with SPEI_FALSE_RECOGNITION. If a false recognition occurs, the application can examine the audio (or even a partial recognition result) to reprocess the recognition or attempt to process the partially recognized text. SAPI does not require that an SR engine send a phrase with the false recognition event. ISpPhrase::GetText retrieves elements from a text phrase. All text recognized is then converted to lower case for the purpose of code generation because C++ is case sensitive and all reserved words are in lower case.

3.3 Code Generator

Code generator is the module that actually generates C++ code from listened words. As a first step, we find a list of reserved words of C++. Now for each reserved word we find words with similar sound. For example “while” have following words with similar sound. "lloyd", "while", "white", "wine" and "voice ".

After this, for each reserved word we developed a separate list structure of words with similar sound in C#. When a user speaks a reserved word that word will be converted to text and this text will be matched to the elements of list structure. If a match is found converted text is replaced with that reserved word. If no match is found text is written as it is. Now if this text is wrong and user wants to remove that word user will speak “incorrect”. A list structure is also maintained for same utterances of “incorrect”. If spoken word is matched with that same utterance then that word is removed.

At the same time if a match is found and that reserved word has special program construct then that program construct is also generated simultaneously.

Input: number sign
Output: #
Input: include
Output: include
Input: iostream
Output :<iostream.h>
Input: void main function
Output: void main(){
}
Input: See out
Ouput: cout<<”
Input: sea in
Output: cin>>
Input: For
Ouput: for
Input: Loop
Ouput: for(; ;){
}
Input: Int I equals to zero;
Output: for(int i=0){
}
Input: I less than ten
Ouput: for (int i = 0; i < 10) {
}
Input: I plus plus
Ouptput: for (int i = 0; i < 10; i++) {
}

This process will continue until user completes his code. When code is complete, User will speak “select all” to select whole code. After selecting code user will speak “copy” and in C++ compiler will speak paste. All code will be pasted here. Now user will have to debug and compile this code in C++ compiler.

Figure.2 shows algorithm for the system.

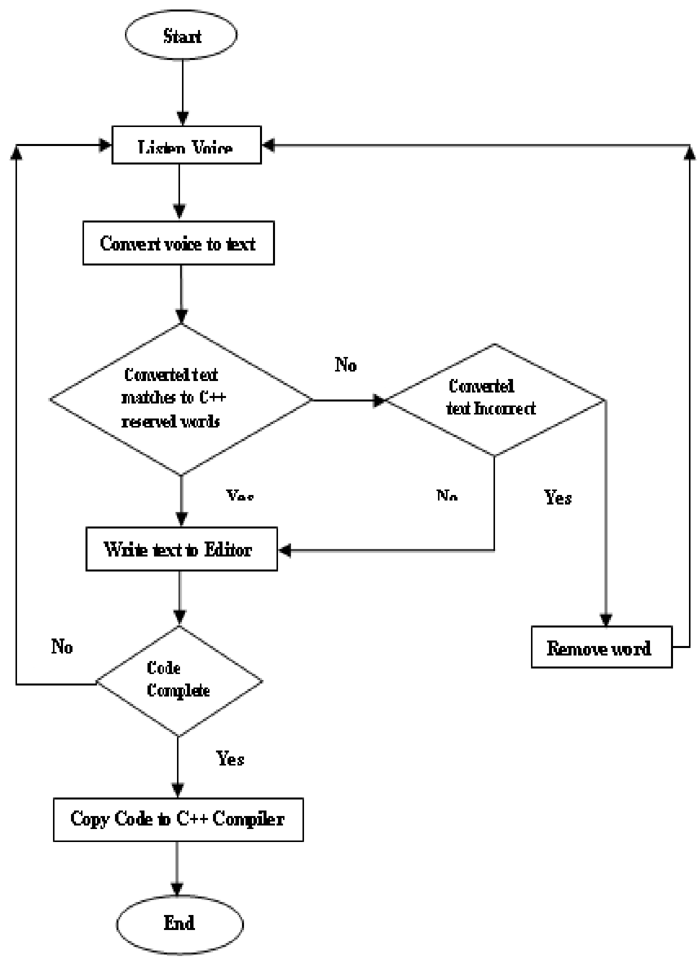


Figure.2 Algorithm

4. COMPARISON

For architectural comparison, we compared our system with various other voice programming systems available in literature e.g. VoiceGrip[8] is a programming tool designed specifically for voice input. The system allows programmers to dictate code using a pseudo code syntax

that is easier to utter than the native code itself. Later, these pseudo code statements can be automatically translated in real time to native code in the appropriate programming language. For example, to type the below C statement, you will simply say:
"if one element is less than second element then"
At first, the statement would get typed literally in pseudo code, but you could later utter a single voice command to automatically translate it to the appropriate native C code.
This approach to code dictation imposes a lower vocal load on the user because it is much shorter to utter. When translating pseudo code to native code in a particular language, VoiceGrip uses a simple deterministic parsing algorithm. At each iteration, it translates a substring starting at the beginning of the remaining pseudo code. It then removes that substring from the pseudo code and proceeds with the next iteration. The process continues until there is no more pseudo code to translate. The utterance is translated in a single pass with no backtracking. This translation of each substring to native code makes translation process too much slow.
SpokenJava [6] system takes the form of a program editor called SPED and associated program analysis framework called Harmonia which are both embedded in the Eclipse IDE. A user begins by speaking some program code in Spoken Java into the editor. Once it has been processed by the voice recognizer, it is analyzed by Harmonia. Harmonia can recognize, handle and support ambiguities through the syntactic phases of program analysis as well as execute semantic analyses to disambiguate the myriad possible interpretations of the input that the first two phases create. When semantic analysis results in several legal options, our programming environment defers to the programmer to choose the appropriate interpretation. Once the interpretations have been deduced, they are translated back
into Java, and written into the editor. SpokenJava is a better system but it is targeted for Java and not for C++.
[9] Consists of adding Java support to VoiceCode. This implementation consisted mainly of adding commands (loop templates, etc.) and their spoken forms to the VoiceCode program. Where possible, I kept the spoken forms for Java consistent with spoken forms in other languages. I built in extremely common commands (println, main method) and set them up to do a lot of automatic typing for the user. Two major limitations of VoiceCode are the complexity of installation and the amount of hand use involved in startup. In order to limit the amount of typing and mouse use required to start VoiceCode, I have created a batch file to start Dragon NaturallySpeaking and VoiceCode.
NaturalJava [1] is a prototype for an intelligent natural-language based user interface for creating, modifying, and examining Java programs. The interface exploits three subsystems. The Sundance natural language processing system accepts English sentences as input and uses information extraction techniques to generate case frames representing program construction and editing directives. A knowledge-based case frame interpreter, PRISM, uses a decision tree to infer program modification operations from the case frames. A Java abstract syntax tree manager, TreeFace, provides the interface that PRISM uses to build and navigate the tree representation of an evolving Java program. The goal of the NaturalJava interface is to allow programmers to write computer programs by expressing each command in natural language. But in our case user will speak native C++ code and not in natural language.

5. CONCLUSION AND FUTURE WORK

Disabled persons or Programmers suffering from repetitive strange injuries will always find it difficult for adapting to software development environments that promote long hours in front of a keyboard. Our work helps make this easier by enabling programmers to use voice recognition. In this paper we presented a robust solution for speech based programming in C++. Implementation mainly consists of finding words that have similar sound to each reserved word of C++ programming language e.g "for" have similar sound to "far", "four" and "thought". All code spoken by the user will be written in an editor. If some text is written that does not match to user intentions user will speak "incorrect" and that word will be removed. Special program constructs (e.g function structure) are also created for the sake of user and thus freeing user form the headache of remembering syntax. Our system is very suitable for disabled persons or persons suffering from repetitive strange injuries
For future development, this approach can be extended to all textual programming languages and also to visual programming languages. Similar approach can also be helpful for query writing.

Acknowledgments

We would like to acknowledge our chairman Dr. Tabassam Nawaz and our family members.

References

[1] Price, David, Riloff, Ellen, Zachary, Joseph and Harvey, Brandon. NaturalJava: A Natural Language Interface for Programming in Java. In the Proceedings of the International Conference on Intelligent User Interfaces, January 2000.
[2] Arnold, Stephen, Mark, Leo and Goldthwaite, John. Programming By Voice, Vocal Programming. In the Proceedings of the ACM 2000 Conference on Assistive Technologies. November 2000.
[3] Desilets, Alain. VoiceGrip: A Tool for Programming by Voice. International Journal of Speech Technology, 4(2): 103-116. June 2001.
[4] Yasuhrio Minami and Sadoki Farui "A large vocabulary continues speech recognition algorithm and its application to a multi-modal telephone directory system" NTT Human Interface Laboratries, Tokyo, Japan

[5] George C. Demetriou and Eric S. Atwell "Semantics in Speech Recognition and Understanding: A Survey", Artificial Intelligence Division, School of Computer Studies, University of Leeds
[6] Begel, A. Programming By Voice: A Domain-specific Application of Speech Recognition. AVIOS Speech Technology Symposium–SpeechTek West (2005).
[7] Douglas A. Reynolds "Automatic Speaker Recognition: Current Approaches and Future Trends" MIT Lincoln Laboratory, Lexington, MA USA.
[8] A. DESILETS VoiceGrip: A Tool for Programming-by-Voice INTERNATIONAL JOURNAL OF SPEECH TECHNOLOGY 4, 103–116, 2001
[9] Java Programming Using Voice Input: Christine Masuoka University of Maryland, 2008
[10] http://www.c-sharpcorner.com/Forums/Thread/75127/speech-recognition.aspx
[11] http://www.dreamincode.net/forums/topic/75990-convert-speech-to-text-and-vice-versa-in-c%23net/
[12] http://msdn.microsoft.com/en-us/library/ms720151(v=VS.85).aspx#API_Speech_Recognition

Mubbashir Ayub received his B.Sc. degree in Software Engineering from University of Engineering & Technology Taxila, Pakistan in 2007. Currently he is enrolled in M.Sc degree of Software Engineering and also serving as Lecturer in the same University in the department of Software Engineering. His areas of interests are Speech Recognition, Software Design & Architecture and Software Testing.

Engr. Raja M. Asjad Saleem did his Bachelors in Software Engineering from University of Engineering & Technology Taxila and Masters in Computer Engineering from the same in 2005 and 2009 respectively. His areas of interest are Software Specification and Design, Information Systems and Engineering Management. Engr. Raja has over 05 years teaching and research experience in University of Engineering and Technology Taxila. He has worked in different domains of technology including Computer Networks, Wireless Security. He is planning to pursue his PhD studies from a foreign university.

Handwriting and Hand Drawing Velocity Modeling by Superposing Beta Impulses and Continuous Training Component

**Houcine Boubaker , Aymen Chaabouni , Najiba Tagougui ,
Monji Kherallah , Adel M. Alimi**

REGIM : Research Groups on Intelligent Machines, University of Sfax, National Engineering School of Sfax (ENIS), BP 1173, Sfax, 3038, Tunisia

Abstract

We present in this paper a new strategy of handwriting or hand drawing velocity modeling Based on the Beta theory. The introduced approach aims to improve the interpretability of the dynamic profile model, reduce the data redundancy, and ameliorate the features accuracy. Indeed, we showed that the curvilinear velocity of handwritten or hand drawn trajectory can be rebuilt by superposing two components; consecutive Beta impulses representing its amplitude alternation imposed by the trajectory curvature variation and a velocity gain part of persistent pen carrying called "continuous training component" interpreting the learning level of the hand drawing faculty and the control of neuromuscular pulses synchronization. The proposed strategy was validated by the reduction of the error of curvilinear velocity fitting and the improvement of the recognition rate of Arabic handwriting characters represented by its model features vector.

Keywords: **Online hand drawing – Beta theory – velocity profile modeling – beta impulse – continuous training component.**

1. Introduction

The velocity modeling is a useful stage in various on-line hand drawn trajectory analyses and recognition process as: handwriting and hand drawn symbols recognition, writer identification, signature authenticity verification, biomechanical system diagnostic, … Different approaches are addressed for the hand drawing velocity modeling. From the oscillatory model of Hollerbach [8], to the Beta elliptic model of Bezine [5] and Kherallah et al [3], preceded by the delta-lognormal model of Plamondon et al [6, 7] and the Beta model of Alimi et al [1, 2], the hand movement drawing speed was always approximated by an association of bell shaped function. We propose in this paper a new strategy of handwriting and hand drawing curvilinear speed modeling based on the Beta approach. It suggests that for a planed hand movement trajectory, the effect of the overlapped neuromuscular subsystems actions on the velocity profile appear as an optimal arrangement to ensure a continuous component to the movement which is superposed with another impulsive component to meet the trajectory curvature variation.

The optimization level result as a compromise between increasing the average level of the continuous training component and the constraint of drawing precision.

Thus, to improve the interpretability of the dynamic profile model, the proposed approach decomposes the velocity profile in two superposed components: beta impulses to represent the trajectory curvature variation and continuous training component to describe the control and training level of the hand drawing faculty.

In order to validate the proposed strategy, we tested its pertinence for the velocity profile modeling of on-line handwriting character and signature by considering the error of velocity profile rebuilding and their rate of recognition.

In the second section of this article, we study the dynamics characteristics of the hand drawing movement. The third section presents the classic overlapped Beta impulse approach for hand movement velocity modeling. Then we introduce the strategy of dynamic profile modeling by Beta function and continuous training component. Finally, we conclude with the result of the model pertinence tests and perspectives.

2. Velocity Profile Modeling by Beta Impulse Overlapping

A handwriting or hand drawing movement, which is generated by neuromuscular excitations, is characterized by its velocity and trajectory profiles. Based on kinematics studies, the Beta model is proposed as modeling tool for the dynamic data of on-line hand drawing movements [1, 2, 5, 3, 7, 9, 14, 15].

3.1 Segmentation of the handwritten trajectory

The hand drawing movement, as any other driving process, is programmed partially in advance. The movements are represented and organized in the velocity fields [6, 7, 4]. In this context, trajectory model is the result of the activation of N neuromuscular subsystems which are characterized by a standardized impulse response. According to the works of Alimi et al. [1, 2], the response of global impulse converges with a Beta curve. Curvilinear velocity $V_\sigma(t)$, calculated by the equation (2), represents the resulting response to the finished impulses. It is smoothed by a second order derivative filters [5, 3] :

$$V_\sigma(t) = \sqrt{\left(\frac{dx(t)}{dt}\right)^2 + \left(\frac{dy(t)}{dt}\right)^2} \qquad (2)$$

The trajectory of the handwriting is segmented in simple movements that called strokes. The number of strokes of one script is determined by an inspection of extremums [1, 7, 5, 3] to know, the local extremas of the horizontal or vertical direction [4], the local extremas of the curvilinear velocity signal of handwriting, and its inflexion points [3] (see figure 2).

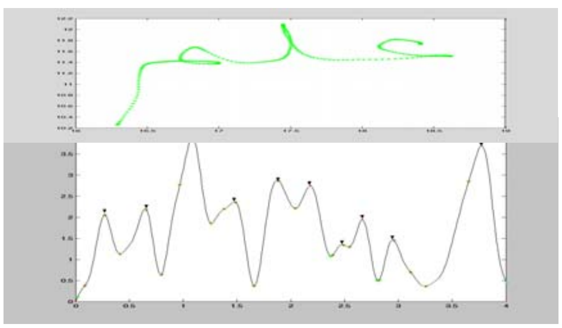


Fig. 2 Detection of the extremum points.

The inflexion points given by acceleration signal are considered and allotted to forms of Beta functions rides in the profile velocity.

3.2 Modeling of the velocity profile

The curvilinear velocity of each stroke obeys to "Beta" approch. Thus, the generation of a model for the trajectory is the algebraic result of the addition of the velocity profiles of the successive strokes (see eq 3).

$$V_\sigma(t) = \sum_{i=1}^{n} V_i(t - t_{0i}) \qquad (3)$$

Consequently, the complete velocity profile,which is generated by the neuromuscular system, is described by the following Beta model :

$$V_r(t) = \sum_{i=1}^{n} K.\beta(t,q,p,t_0,t_1) \qquad (4)$$

With :

$$\beta(t,q,p,t_0,t_1) = \begin{cases} \left(\frac{t-t_0}{t_C-t_0}\right)^p \cdot \left(\frac{t_1-t}{t_1-t_C}\right)^q & \text{if } t \in [t_0, t_1] \\ 0 & \text{elsewhere} \end{cases} \qquad (5)$$

Where t_0 is the starting time of Beta function,
t_c is the instant when the curvilinear velocity reaches the amplitude of the inflexion point,
t_1 is the ending time of Beta function checking $t_0 < t_1 \in IR$, and p , q are intermediate parameters, which have an influence on the symmetry of Beta shape and verifying :

$$\frac{t_C - t_0}{t_1 - t_C} = \frac{p}{q} \qquad (6)$$

The shape of a symmetrical Beta signal is given by figure 3 . The parameter K is the amplitude of the beta signal.

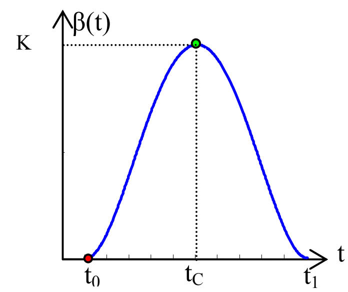


Fig. 3 Shape of a symmetrical Beta impulse function for p = q = 2.5

The curvilinear velocity results as the superposition of the neuromuscular finished actions with the impulsive character which is modeled by overlapping Beta impulses in the course of time (see Figures 4a and 4b).

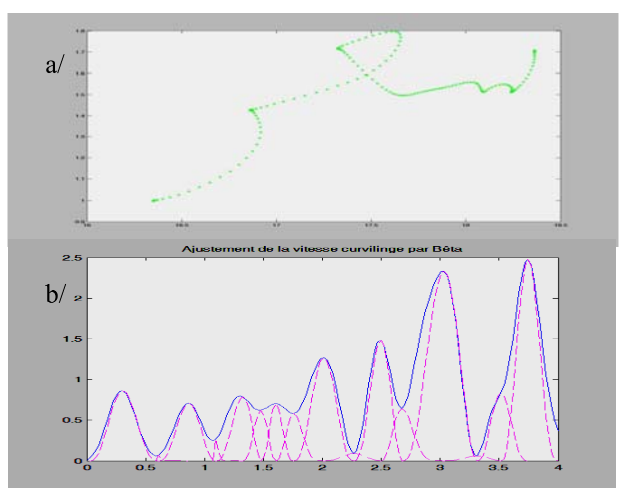


Fig. 4 Velocity signal modeling by overlapping Beta functions

4. News strategy of velocity profile Modeling

To enhance the interpretability of the Beta model, we introduce in this section a new strategy for velocity profile modeling by superposing successive beta impulses to a persistent component of pen carrying called continuous training component.

4.1 Principle

Observing the evolution of handwriting or hand drawing executed by young children (4 to 8 years), we note that they generally actuate their pencils in an impulsive and discontinuous mode. Their drawing or script displays acute forms, discontinuities and distortions [11, 12]. Theirs pencils velocity is cancelled at several times during a continuous line drawing. Later learned and trained to perform hand drawing (more then 9 years), their drawn trajectory becomes more continuous, cursive, faster and less acute (see Figures 5). Indeed, their control of the acceleration and braking actions becomes precise what enables them to avoid the cancellation of the trajectory velocity during a continuous line drawing by maintaining a not null component of drag [11, 12].
However, the hand drawing velocity variation keeps always a relative impulsive character which interprets the intrinsic curvature radius variation of the executed trajectory. These successive impulsions are superposed to the component of continuous drag developed by training.

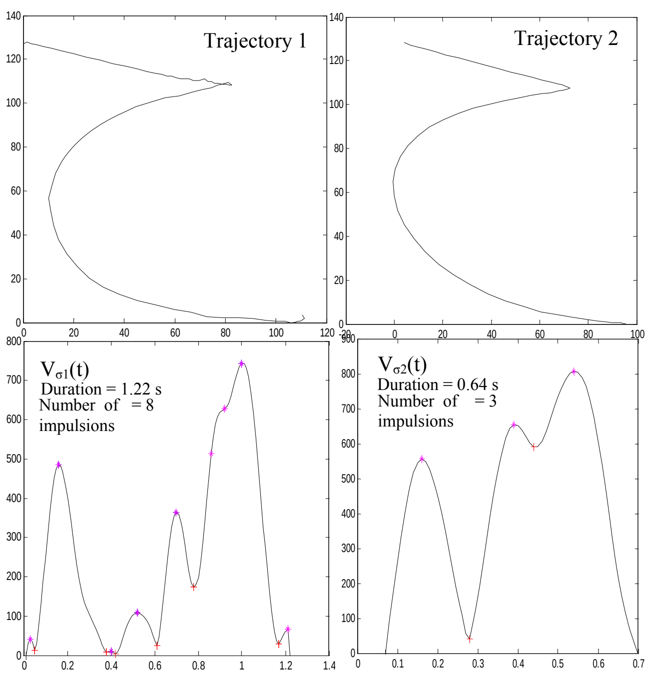


Fig. 5 samples of handwriting trajectory and corresponding velocity profile executed by two schoolchildren of respectively 8 and 13 years

4.2 Velocity profile modeling by superposing Beta impulses and continuous training component

We decompose the time axis of the profile velocity into intervals which represent cycles of acceleration, deceleration and braking. Each time interval $T=[t_0, t_1]$ is limited by a successive local minimums or double inflexion points of velocity: $V_i=V_\sigma(t_0)$ and $V_f = V_\sigma(t_1)$. During each interval, the curvilinear velocity can be fictitiously divided into two components:

- An **imp**ulsive component: $V_{Imp}(t)$

  It is a velocity impulsion during the interval T with finished energy, engendered by a cycle of acceleration, deceleration and braking. It can be modeled by a Beta function : (6)

$$V_{Imp}(t) = K \cdot \left(\frac{t - t_0}{t_C - t_0}\right)^p \cdot \left(\frac{t_1 - t}{t_1 - t_C}\right)^q$$

- A continuous **tra**ining component: $V_{Tra}(t)$

  It engenders the energy which allows the continuous passage (with a not null velocity) from a trajectory segment to another separated by a local minimum of curvature radius. Its variation must have the most monotonous and softest character in order to reserve the velocity impulsive character to $V_{Imp}(t)$ component.

Thus, $V_{Tra}(t)$ represents the initial velocity gain V_i to which we add the algebraic effect of a supplementary energy (of acceleration or braking) added to the finished impulse $V_{Imp}(t)$ in order to ensure the assymetry of the curvature radius variation from $R_i = R(t_0)$ to $R_f = R(t_1)$ (see Figure 6).

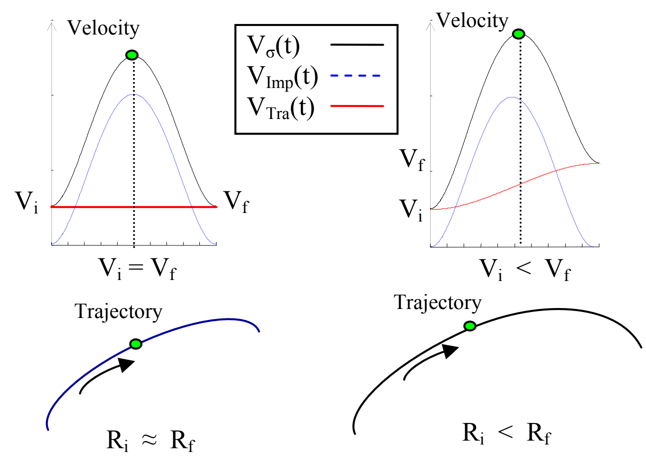


Fig. 6 Correspondence between asymmetry of curvature radius and velocity variations

The variation according to the time of the continuous training component $V_{Tra}(t)$ is given by a monotonous polynomial function of third degree :

$$V_{Tra}(t) = a \cdot \left[\frac{(t-t_0)^3}{3} - \frac{(t_1-t_0)\cdot(t-t_0)^2}{2} \right] + V_i \quad (7)$$

where $a = -6 \cdot \dfrac{V_f - V_i}{(t_1 - t_0)^3}$

The reconstituted curvilinear speed of tracing is obtained by the sum of its impulsive component with the component of continues drag :

$$V_R(t) = V_{Imp}(t) + V_{Tra}(t) \quad (8)$$

(see Figure 7)

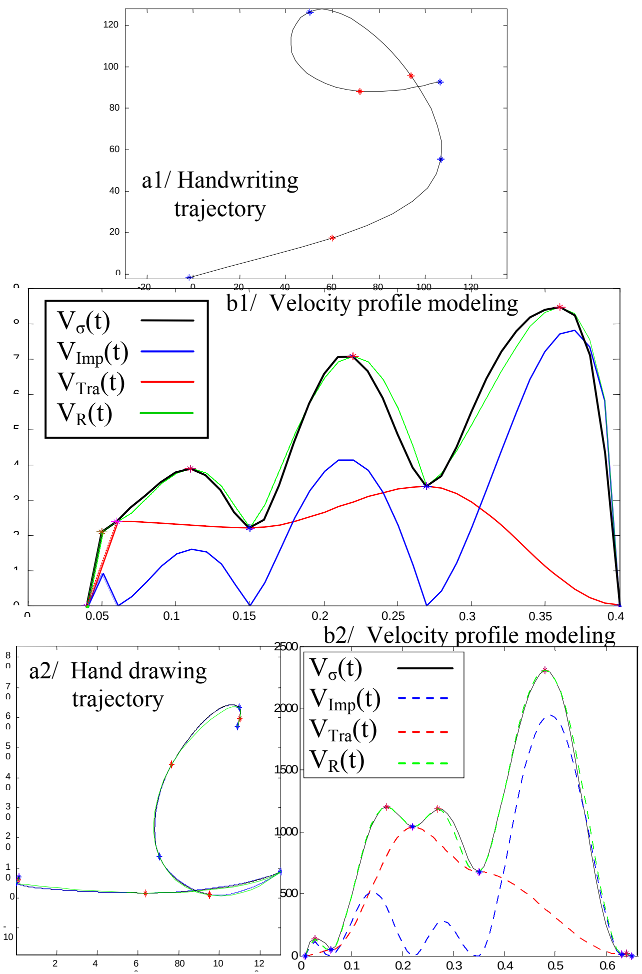


Fig. 7 Examples of velocity profile modeling by superposing successive Beta impulses and continuous training component

The parameters K, and t_c of a given beta impulse component $V_{Imp}(t)$ are given from the original velocity profile. The t_c moment corresponds to the local maximum velocity: $\dfrac{d\, V_R(t_c)}{dt} = 0$. This leads to the following relation between p, q and t_c :

$$\frac{p}{t_C - t_O} - \frac{K.q}{t_1 - t_C} = \left[\frac{a}{2}.(t_1 - t_C).(t_C - t_O) \right] \quad (9)$$

4.3 Simplified variant of the Beta modeling approach

This new strategy adopts a simplified representation of the real effect of neuromuscular impulses overlapping in its various modes of implementation (see figure 8). Indeed, the resulting effect of this overlap is modeled by the superposition of consecutive and finite speed impulses modeled by Beta functions on a continuous training component representing the inferior limit of the curvilinear speed variation envelope.

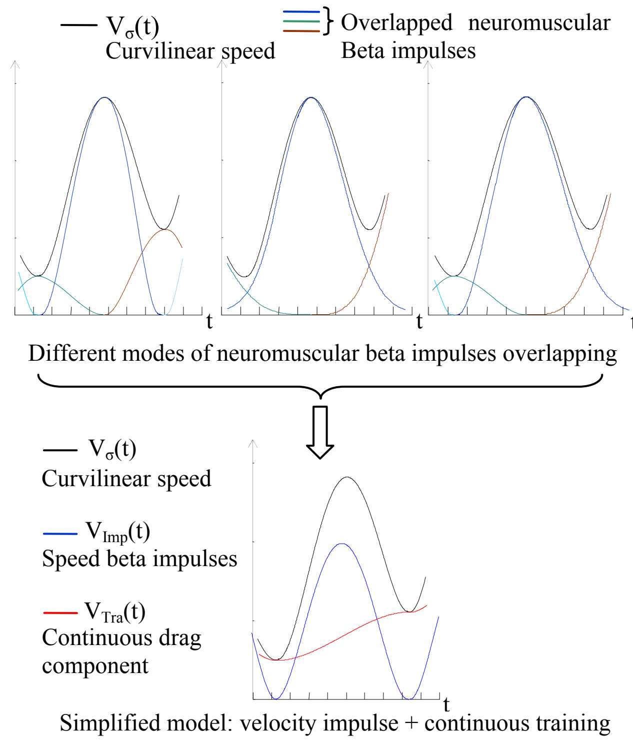


Fig. 8 Strategy of simplification of the velocity Beta modeling approach

5. Evaluation of the Modeling Approach Pertinence

5.1 Reduction of the data load

The new modeling strategy reduces the data redundancy by adopting a simple segmentation approach of not – overlapped velocity Beta strokes. In fact, a hand drawn trajectory with $n = (2 \cdot m) + 1$ successive speed extremums; $m+1$ minimums alternated by m maximum speed, is segmented into m strokes by the new strategy when it is segmented into $(2 \cdot m) - 1$ strokes using the overlapped Beta approach. The decrease in the number of segmented strokes reduces the total number of the considered parameters despite the increment of the features vector size from 4 parameters $[K, \Delta t_1 = (t_1 - t_0), \Delta t_C = (t_C - t_0), p]$ for the overlapped Beta approach to 6 parameters $[K, \Delta t_1 = (t_1 - t_0), \Delta t_C = (t_C - t_0), p, a, V_i]$ with the new strategy where the added couple of parameters $[a, V_i]$ model the continuous training (drive) component.

Thus, to model the velocity profile of the drawing of a wavy line of 7 peaks or the handwriting of the word "mini" succeeding each one 15 minimums alternated with 14 local maximums of speed, Beta overlapped approach would use:
$((2 \times 14) - 1)\, strokes \times 4\ parameters = 108\ parameters$
while the strategy of Beta impulse and continuous training component would need :
$14\ strokes \times 6\ parameters = 84\ parameters$

However this parameters load easing in no way affects the modeling accuracy. In fact, the proposed strategy makes it possible to simplify the problem of calculating the shape parameters *p* and *q* of the velocity Beta pulses avoiding the interdependence of this parameters for all the generated impulses due to their overlap that will be modeled locally by the continuous training component $V_{Tra}(t)$. Thus, determining the parameters of the speed Beta impulse $V_{Imp}(t)$ correspondent to a specified time interval depends only on the variation of the curvilinear speed $V_\sigma(t)$ during this interval and that of the continuous training (drive) component $V_{Tra}(t)$.

5.2 Evaluation of the Accuracy of velocity profile fitting

The reduction in the number of unknowns allows more stability for the regression system computing the values of the parameters *p* and *q* and then better precision on their estimation which leads to reduces the velocity profile reconstruction error. In fact, tests are conducted on a set composed of 5000 samples of isolated Arabic characters from the LMCA database [4], 500 signatures (see in Figure 9 an example of signature dynamic profile modeling) and 1000 handwritten symbols, to study the accuracy of the new strategy of hand movement velocity profile modeling. For each sample we calculate the rate in percent of the fitting error of curvilinear velocity profile at each point of the trajectory:

$$Err\_V(t) = \frac{|V_\sigma(t) - V_R(t)|}{V_\sigma(t)} \times 100 \qquad (10)$$

where $V_R(t) = (V_{Imp}(t) + V_{Tra}(t)) \qquad (11)$

Then the average error over the whole trajectory of the sample:

$$Ave\_Err\_V = \frac{\int_{t_1}^{t_2} Err\_V(t) \cdot dt}{|t_2 - t_1|} \qquad (12)$$

The following table compares the results of the average error of dynamic profile reconstruction obtained for different types of hand - drawn graphics respectively by the simplified strategy (superposition of Beta impulses and continuous training component) and the overlapped Beta impulses model taking the parameter q modeling the shape of the decay phase as a constant q = 2 :

Table 1: Dynamic profile reconstruction error rates obtained by the overlapped and the simplified Beta strategies

Strategy of velocity profile modeling	***Ave_Err_V***		
	Isolated characters	signatures	Symbols
Overlapped Beta impulses with approximation *q* = constant	22.8 %	23.4 %	20.6 %
Superposition of Beta impulses and continuous training component	9.5 %	10.3 %	8.7 %

The results show an improvement of the accuracy of velocity profile fitting by adopting the simplified strategy of the Beta modeling.

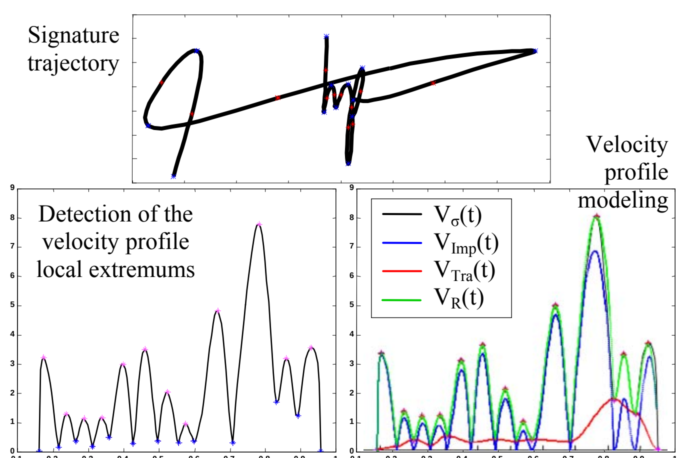


Fig. 9 Examples of Signature velocity profile modeling by superposing Beta impulses and continuous training component

5.3 Evaluation of the model discrimination power

Both strategies of dynamic profile modeling were also tested for recognition of isolated Arabic characters. These latter are grouped into categories defined by the number of strokes of the characters samples that include. We have listed eight categories for the overlapped Beta strategy from 5 to 12 strokes characters and only five categories for the simplified strategy going from 3 to 7 strokes characters.

The overall recognition system consists of a number of neural network subsystems, each of which is associated with a category of stroke number. Each subsystem is composed by neural networks type OCON (One Class One Network). The number of OCON in each subsystem is equal to the number of existing classes in the category of features. For example, the subsystem of 3 strokes is composed by four OCON, one for each following character label : letter ' ا ' 'alif', letter ' ل ' 'lam', letter ' ن ' 'noun' and letter ' ر ' 'ra' .

The recognition system learning is performed on two-thirds of the 5000 samples composing the set of on-line Arabic handwriting characters of the LMCA database [4]. The remainder third of the database is used as a test set [10, 13].

After its trajectory dynamic profile modeling, each sample is first assigned to the subsystem of category that corresponds to the number of Beta strokes composing it. Then, the activated recognition subsystem presents the features vector of the tested sample to the different OCON associated to its category. Finally, the tested sample is assigned to the character label corresponding to the OCON that maximizes its recognition rate. The following tables show the results of recognition tests for the character class: 'ا ' 'Alif', 'ن' 'Noun', ' هـ ' 'Haa', 'ص' 'Sad', 'و ' 'Waw', 'ف' 'Fa', and 'س' 'sin', obtained using respectively the simplified strategy superposing Beta impulse and training component and the strategy of Beta impulses overlapping when fixing the parameter q as a constant equal to 2:

Table 2: Results of recognition tests obtained using the simplified strategy superposing Beta impulse and continuous training component

Category	Handwritten character	Recognition rate in (%)
3 Strokes	Alif	99.90
	Noun	82.13
4 Strokes	Haa	100
	Noun	86.45
5 Strokes	Sad	88.41
	Haa	87.50
	Waw	89.57
6 Strokes	Fa	88.93
7 Strokes	Fa	84.38
	Sin	84.79

Table 3: Results of recognition tests obtained using the strategy of Beta impulses overlapping (q=2)

Category	Handwritten character	Recognition rate in (%)
5 Strokes	Alif	97.78
	Noun	86,13
6 Strokes	Haa	89.12
	Noun	82.22
	Alif	83.85
	Waw	83.80
7 Strokes	Haa	87.09
	Noun	88.22
	Waw	74.62
8 Strokes	Haa	88.93
	Fa	78.04
9 Strokes	Sad	82.87
	Fa	80.60
10 Strokes	Sad	81.67
	Fa	75.34
11 Strokes	Sad	84.21
12 Strokes	Sin	82.85

The results show an improvement in the average recognition rate obtained with the implementation of the simplified strategy of the Beta model compared to that obtained by the overlapped Beta impulses approach.

6. Conclusions

We have presented in this work a new strategy of handwriting and hand drawing dynamic profile modeling based on the Beta model theory. The introduced strategy decomposes the curvilinear velocity into two components: a component of pen continuous training (drag) that simplifies the representation of the dynamic effect of neuromuscular impulses overlapping, and a second component of sequenced and not overlapped Beta impulses that represent the alternate variation of the curvilinear velocity amplitude. The proposed strategy is a simplified variant of the Beta modeling approach that reduce the data redundancy and improve the model interpretability. Its validation is proved by the reduction of the error of curvilinear velocity fitting and the amelioration of the recognition rate of Arabic handwriting characters.

Acknowledgments

The authors would like to acknowledge the financial support of this work by grants from the General Direction of Scientific Research and Technological Renovation (DGRST), Tunisia, under the ARUB program 01/UR/11/02.

References

[1] A. M. Alimi and R. Plamondon, "Analysis of the Parameter Dependence Handwriting Models Generation one Movement Characteristics", in C. Faure, P. Keuss, G Lorette, and A. Vinter (eds.) Advances in Handwriting and Drawing, Europia, Paris, 1994, pp. 363-378.

[2] M. A. Alimi, "The System Beta: Toward has Change in Our Use of Neuro-Fuzzy Systems", International Newspaper of Management, Invited Paper, No. June, 2002, pp. 15-19.

[3] M. Kherallah, L. Haddad, A. M. Alimi, & A. Mitiche, "Online handwritten digit recognition based on trajectory and velocity modeling", Pattern Recognition Letters, volume 29, 2008, pp. 580–594.

[4] P. Morasso and F. A. Mussa Ivaldi, "Trajectory formation and handwriting: A computational model", Biological Cybernetics , volume 45 , 1982, pp. 131–142.

[5] H. Bezine, M.A. Alimi, and N. Derbel, "Year Explanation for the feature of Handwriting Trajectory Movement Controlled by Beta Elliptic Model", Proc. 7th Int. Conf. one Document Analysis and Recognition ICDAR' 03, Edinburgh, Scotland, the U.K., Aug., 2003, pp. 1228-1232.

[6] R. Plamondon, W. Guerfali, "The generation of handwriting with delta-lognormal synergies" , Biological Cybernetics C, volume 78, 1998, pp. 119–132.

[7] R. Plamondon, A. M. Alimi, P. Yergeau, and F. Leclerc, "Modeling Velocity Profile of Rapid Movements with Comparative Study", Biological Cybernetics, vol. 69, 1993, pp. 119-128.

[8] J. Hollerbach and T. Flash, "Dynamic Interactions between Limb Segments during lunar ARM Movements", Biological Cybernetics, vol.44, 1982, pp. 67-77.

[9] S. Njah, M. Ltaief, H. Bezine and A. M. Alimi, "The PerTOHS Theory for On-Line Handwriting Segmentation", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 3, September 2012, pp 142 – 151.

[10] I. Muthumani and C. R. Uma Kumari, "Online Character Recognition of Handwritten Cursive Script", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 2, May 2012, pp 352 – 354.

[11] H. Boubaker, M. Kherallah and A. M. Alimi, "New Strategy for the On-Line Handwriting Modelling,", in proceeding of the 9th International Conference on Document Analysis and Recognition ICDAR, September 2007, Curitiba, Paraná, Brazil, Vol 2, pp.1233-1247.

[12] H. Boubaker, M. Kherallah, and A. M. Alimi. "Une nouvelle stratégie pour l'extraction des caractéristiques de l'écriture manuscrite en ligne", In Actes du 9ème Colloque International Francophone sur l'Ecrit et le Document CIFED, , 2006, September, pp. 132-138

[13] Z. Kahar, M. Ibrayim, D. Tursun and A. Hamdulla, "Peripheral Contour Feature Based On-line Handwritten Uyghur Character Recognition", IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 2, No 2, Marsh 2013, pp 273 – 278.

[14] N. Tagougui, M. Kherallah, and A. M. Alimi, "Online Arabic handwriting recognition: a survey," International Journal on Document Analysis and Recognition (IJDAR), 2012, pp. 1-18.

[15] A. Chaabouni, H. Boubaker, M. Kherallah, H. El-Abed, and A. M. Alimi, "Static and Dynamic Features for Writer Identification Based on Multi-fractals", International Arabic Journal on Information Technologies (IAJIT), 2013.

Houcine Boubaker was born in Kalaat El Andalous (Tunisia) in 1973. He graduated in Electrical Engineering in 1995, obtained a master degree in Systems Analyses and Digital Signal Processing in 1997. He is a researcher and a PhD student in Electrical & Computer Engineering at the University of Sfax. His research interest includes trajectory modeling and applications of intelligent methods to pattern recognition. He focuses his research on drawing, Arabic handwriting and arm – hand movements modeling and Analyses. He is an IEEE student member and affiliate to the Research Group on Intelligent Machines laboratory (REGIM).

Aymen Chaabouni was born in Sidi Bouzid (Tunisia) in 1981. He obtained the master degree in computer science from the university of Avignon and the Vaucluse, French in 2006. Currently he is a researcher and a PhD student in Computer Engineering at the University of Sfax. His main research interest is in the area of handwriting recognition, writer identification and signature verification. He is an IEEE student member and affiliate to the Research Group on Intelligent Machines laboratory (REGIM).

Najiba Tagougui was born in Sfax (Tunisia) in 1982. She graduated in Computer Sciences in 2005, obtained a master degree in News technologies of dedicated computer systems in 2007. She is now a PhD student in Computer Systems Engineering at the University of Sfax. His research interest includes applications of intelligent methods to pattern recognition. She focuses her research on intelligent pattern recognition especially Arabic Handwriting Recognition. She is an IEEE student member and affiliate to the Research Group on Intelligent Machines laboratory (REGIM).

Monji Kherallah was born in Sfax (Tunisia) in 1963. He graduated in Electrical Engineering 1989, obtained a PhD in Electrical Engineering in 2008. He is now a professor in Electrical & Computer Engineering at the University of Sfax. His research interest includes applications of intelligent methods to pattern recognition and industrial processes. He focuses his research on intelligent pattern recognition especially Arabic Handwriting Recognition. He is member of the editorial board of "Pattern Recognition Letters". He was a member of the organization committee of the International Conference on Machine Intelligence ACIDCA-ICMI'2005.He is an IEEE member.

Adel M. Alimi was born in Sfax (Tunisia) in 1966. He graduated in Electrical Engineering 1990, obtained a PhD and then an HDR both in Electrical & Computer Engineering in 1995 and 2000 respectively. He is now professor in Electrical & Computer Engineering at the University of Sfax. His research interest includes applications of intelligent methods (neural networks, fuzzy logic, evolutionary algorithms) to pattern recognition, robotic systems, vision systems, and industrial processes. He focuses his research on intelligent pattern recognition, learning, analysis and intelligent control of large scale complex systems. He is associate editor and member of the editorial board of many international scientific journals. He was guest editor of several special issues of international journals (e.g. Fuzzy Sets & Systems, Soft Computing, Journal of Decision Systems, Integrated Computer Aided Engineering, Systems Analysis Modeling and Simulations). He is an IEEE senior member.

Adapting IS Strategic Planning methodology to define Mobile Strategy

Mouhsine LAKHDISSI1, Rachid EL HABI2, Alami SEMMA3, Bouchaib BOUNABAT4

1 Mohammed V Souissi University
Al Qualsadi Laboratory, ENSIAS, - Rabat, Morocco

2 Faculté des Sciences et Techniques, Université Hassan 1er,
Settat, Morocco

3 Faculté des Sciences et Techniques, Université Hassan 1er,
Settat, Morocco

4 Mohammed V Souissi University
Al Qualsadi Laboratory, ENSIAS, - Rabat, Morocco

Abstract

The nature of mobile initiative in a company make them usually opportunistic and pragmatic. This leads inevitably to some incoherence and inconsistence in companies' passage to digital and mobile world. Mobile strategic planning is one solution to make mobile initiatives adhere to the global strategy and respect a more rigorous plan. We introduce in this paper a methodology for mobile strategic planning is based on Enterprise Architecture frameworks and our work on strategic planning SI. This methodology is intended to be light, pragmatic, iterative and incremental, so to be tailored to mobile applications and projects. We furthermore defined several methodological tools and techniques to better support the methodology.

Keywords: *Mobile strategy, IS Strategic Planning, Enterprise Architecture, Mobile usage, Mobile context, Mobile capabilities*

1. Introduction

In few years, new mobile devices, smartphones and tablets, have upset habits of hundreds of millions users around the world. They evolved from a 'simple' device to make a phone call or to send a short message, to appliances, as powerful as PCs, providing new applications like web browsing, mobile commerce transactions or access to a dashboard whose data is located in the enterprise information system.

Unlike traditional PC applications, mobile applications are distinguished by the intimacy characteristic and the integration of the user's contextual information.
Aware of the challenges of mobility, businesses and governments began to offer, more and more, new mobile services to their customers, prospects and partners through innovative solutions that combine responses to the needs of users, opportunities for value creation with mobile device capabilities (and possibly Information System capabilities). Thus, we see the emergence of new "m-services" such as m-health, m-commerce, m-government, m-marketing, m-ticketing, m-coupon, m-banking, m-finance services or applications more content, information or social oriented.

To plan the introduction and evolution of mobility as a new way of communication and management within the enterprise and as a new way to deal with partners and doing business, managers need a mobile strategic planning methodology tailored to mobile applications and projects.
We propose in this paper a methodology for mobile strategic planning based on Enterprise Architecture frameworks and our experience in information strategic planning. This methodology is intended to be light, pragmatic, iterative and incremental.

The second section will introduce strategic planning in general and particularly in the information systems domains. We will describe specific elements of mobility in the third section. The fourth section is dedicated to a detailed description of the proposed methodology and the fifth section will present the tools and techniques supporting the methodology. And we will conclude with a summary and future work.

2. IS strategic planning

2.1 Strategy and strategic planning

Strategy is defined by Chandler as "The determination of the basic long-term goals and objectives of an enterprise and

the adoption of courses of action and the allocation of resources necessary for carrying out these goals" [6] and by Porter as "The art to build durable and defendable competitive advantage"[7].
One of the most complete definitions was given by [8], "A fundamental framework for an organization to assert its vital continuity, while, at the same time, forcefully facilitating its adaptation to a changing environment."
For most of the definition, strategic planning is focused on three main questions:

- Where we are?
- Where we want to go?
- How to get there?

2.2 IS Strategic Planning methodologies

IS Strategic planning has been defined by [9] as the process of identifying a portfolio of applications/projects that can help an organization achieve its business strategy. Its focus is on defining the IT roadmap in term of key initiatives, projects and transformations to be made on the existing information system with two main intentions:

- How to align information systems with business needs and overall strategy?
- How to use information technology to change and impact the business?

Due to the complexity of today's information systems and the diversity of enterprise's technology approaches, many methods have been defined to structure the ISSP process and techniques have been defined to address some aspects of the discipline. [10] classifies ISSP methods into two categories:

- Impact methods : trying to make It help create a positive impact and drive the change of the business
- Alignment methods : where the main focus is on aligning IT to respond to business needs and to help achieve strategic goals

Among the methods used in IT Strategic Planning we have Critical Success Factors (CSF) [1] which could be considered as an impact and alignment method, Business Systems Planning (BSP) [1], Porter's Value Chain [7], and Scenarios [1]. Methods can be grouped together to constitute a methodology. Methodologies used for ISSP include those of the CCTA (12) and Boar (13).
Many IT vendors and consultancy organizations use proprietary methods and/or methodologies, some of which are adaptations of open source approaches. Examples are Arthur Andersen's Method/1 and Coopers and Lybrand's Summit [9]. It is also well known that organizations often develop their own in-house methodologies, often based on open or proprietary methods or approaches [9].

2.3 Our IS Strategid Planning methodology

In [16] and [17] we introduced a new methodology for IS strategic planning. This methodology is inspired from Enterprise Architecture Frameworks and brings more structure, standardization and industrialization to existing methodologies.

One of the major elements of the methodology is the metamodel that is used to describe the enterprise as a whole and the information system specifically. This metamodel brings a way to link the strategy to the defined target and the roadmap projects through the concept of transformation.

The suggested metamodel is composed of five layers:

- Strategy
- Business
- Information Systems
- Technology
- Strategic planning

All layers are interrelated with static and dynamic element of the three natures: function, structure and content. Every layer is connected with the layer below with a realization link. A process is automated in an application which uses a database and are both deployed in a server. This dependency is fundamental to align the IS with the Business Architecture and the Technology with the IS Architecture. This link allows us also to analyze the gap between layers in term of coverage to make it possible to fill this gap in the strategic plan[16][17].

In the context of mobility, we can consider mobile applications as part of the information system with specific attributes.

We suggest adding mobility attributes to all metamodel objects to make sure we capture mobile requirements, constraints and capabilities.

For example, we can add an attribute mobile channel to Business Processes to indicate whether a process could be invoked from a mobile channel. The same applies to services, applications, information and servers.

This will make it possible to analyze the existing portfolio of processes, applications, information and servers to verify their readiness and maturity for mobile enterprise.

3. Mobile specifics and mobile strategy

3.1 Mobiles usages

Basically, mobile applications, whether native, mobile web or hybrid, are made to meet the needs of users that are primarily:

- To be informed: by searching for information on a search engine, browsing websites, reading news or consulting a directory;
- To communicate: through voice, text or video exchange;
- To save and store documentary or multimedia content;
- To collaborate: through sharing documents and media content or accessing social networks;
- To Learn: through reading an ebook or using an e-learning application;
- To supervise or control remotely: as being notified of special events from an information system or remote video surveillance;
- To access or receive a professional or institutional service as checking a bank balance or to apply to a public administration service;
- To perform e-commerce transaction: through purchasing and making sales with payment features;
- To work remotely by accessing the company resources and the information systems;
- To entertain: through music, games and movies.

These needs, when they are combined with a situation, habits or preferences of usage (information) or user behavior or external events, are called **mobile usage**

3.2 Mobile capabilities

New technological capabilities of mobile devices offer more levers of innovation and opportunities to impact the company's business. Mobile devices are now equipped with a set of capabilities like communication features (Voice, SMS / MMS, Email), Internet access for browsing and cloud services and information system access, input capabilities (keyboard, touch screen, voice recognition, camera) and display (text, image, video) and capabilities related to internal and external sensors (GPS Geolocation, Accelerometer, gyro, proximity ... etc.). These capabilities are classified and summarized in the technical reference model proposed in section 5. We will see in the near future devices equipped with chemical sensors that can be used, for example, to assess the freshness of food and an integrated 3D camera that can be used to control the movements.

.

Mobile solutions have, also, brought new opportunities to improve employees' productivity and efficiency of business processes once the capabilities of the mobile device are combined with external information systems of the companies with and capabilities existing in the cloud. However, companies' information systems must, first, demonstrate an openness and performance to be able to interact with mobile applications while insuring security.

3.3 Mobile Context

The concept of 'context' of the mobile user is very important to understand the value of mobility in the transformation of the company into the digital age. Several definitions have been proposed by [1] concluded that by its own definition "The context is a set of environmental conditions and preferences that either determine the behavior of the application or that involves an interesting application event for the user." Several definitions bring out some classifications such as the one given by Forrester [2] through the following three elements:

- The Situation: which can be characterized by the current time, location, altitude, environmental conditions, speed, etc..
- The Preferences: built by the previous decisions taken by the user and shared in the application or in social networks
- Attitude: through feelings and emotions inferred from user actions

Another definition of [3] distinguishes the computer context, user context and physical context.

Retrieving the context of the user has become easier thanks to the integration, in devices, of new software features and electronic components such as camera, GPS, accelerometer, sensors, NFC (Near Field Communications) …etc.

We believe that the context is none other than "the combination of a particular usage (situation / relationship / behavior) of the user with a device capacity". This definition introduces the link between the environment and the user need and the device capabilities. We can classify the context as a combination of three aspects:

- Informational context: situation and environmental conditions
- Organizational or structural context: relationships, contacts, networks
- Behavioral context: attitude, preferences and reactions3. Tables, Figures and Equations

3.4 Mobile strategy definition

In the last three years, mobile has become an important topic in business. Beyond its public use, we hear more and more of the mobile terminal as the future (Post-PC era, as a privileged communication channel with external stakeholders (customers, suppliers, partners) and internal stakeholders (employees, managers, shareholders) or as a tool to support the company's business processes.

These changes imply a need to define the company's mobile strategy in line with its overall strategy while taking into consideration the characteristics and constraints of mobiles environments.

The classical approach is to consider mobile applications as part of the Information System and integrate them into the process of IS strategic planning. This approach may be appropriate when the mobile is none other than an access channel to the IS and will apply more for alignment purposes than in the impact logic. The other disadvantage of this approach is that it does not take into account the specificities of mobile development, mobile projects and mobile use.

4. Mobile Strategic Planning Methodology

4.1 Overview

Following on our work of defining a methodology for IS strategic planning described in [16] and [17], we took advantage of the proposed approach that is based on the Enterprise Architecture to define a specific methodology the mobile with an adapted metamodel and taking into account the following elements:

- A lightweight approach adapted to the time constraint usually related to the definition of a mobile strategy
- An iterative and incremental approach to take into account the progressive character and the maturation process of a company to move to mobile. We also plan as a future work to define a maturity model for enterprise mobility.
- An approach that is inspired from a global Framework, which is the TOGAF framework [7] which allows to enrich the Framework and extend the methodology if necessary.
- A pragmatic approach supported by methodological techniques as will be presented in the next section.

The proposed methodology called VERGA consists of five phases:

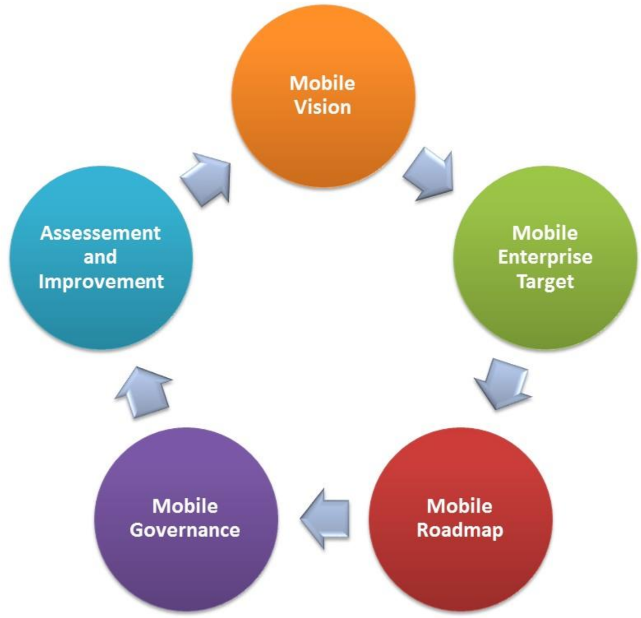


Fig. 1. Mobile Stratgic Planning methodology : VERGA

- **Mobile Vision:** The objective in this phase is to define the strategic objectives of mobility and the Mobile Vision aligned with the strategic business and IT objectives of the company.
- **Mobile Enterprise Target:** In this phase we define the target processes and Information Systems to support the vision.
- **Mobile Roadmap:** This phase aim at defining the action plans necessary to move from the current state to the mobile enterprise target
- **Mobile Governance**: This phase make it possible to set up the organization, processes and tools, and other activities necessary to deploy, manage and maintain mobile initiatives
- **Assessement and Improvement**: To ensure continuous improvement of mobility, we need to assess regularly the needs and capabilities and make evolve the architecture while monitoring KPIs and usage.

4.2 Detailed description

The following table describes more in detail these phases :

Phase	Activities
Mobile Vision	As-is situation evaluation Maturity Assessement Defining the mobile vision in terms of : - strategic objectives, - business requirements, - target users and devices

	- users' needs, - Mobile value chain Defining the existing capabilities and resources : - Business capabilities - Mobile terminal capabilities - IS capabilities - Cloud capabilities Mobile principles and guidelines Benchmark with similar organizations KPIs to assess mobile initiatives
Mobile Enterprise Target	This phase describes the target situation pursued by the enterprise based on the vision defined earlier : User experience: Define principles of user experience and ergonomics that are coherent throughout the enterprise. This should be based on element of the vision (target users, usages and mobile principles) Business View: identify business processes where mobility could have a benefit. Sometimes the mobile integration could imply the creation of new business activities or organization roles. The basic tool for this task is the detailed view of the Mobile Value Chain. IS View: The aim here is to transform the need for automation in one or more process where mobility is introduced in one or more applications using databases and flows. Tools that can be used are: - IS Mobile mapping - Mobile application reference model Technological View: identify the types of mobile devices and technologies to use. The tool can be used is the "Mobile Technical Reference Model". It is also necessary here to identify new needs in terms of hardware and software infrastructure to cope with security, performance and distribution challenges introduced by mobile applications. For example, it is sometimes necessary to acquire Mobile Device Management (MDM) software
Mobile Roadmap	Short-term action plan - Opportunities solution - Capitalizing on the existing IS and especially web applications - Preparation of existing IS application (services exposition) Projects Plan - Detailed mobile projects (initiatives) - Prioritization - Dependencies between mobile projects

	- Dependencies with IT projects
Mobile Gouvernance	Definition of actors, roles and processes Adding mobility requirement and constraints to projects management processes Definition of governance of mobile security Mobile developement industrialization Organizational change management Communication Plan
Assessement and Improvement	Tracking and monitoring of mobile use Tracking KPIs Maintenance and evolution of mobile solutions

5. Methodological tools and techniques

The activities described above should be supported by some tools and techniques. Below, we will outline the « Mobile Vision Cube », a tool to set the Mobile Enterprise guidelines in the "Mobile Enterprise Vision" phase. Some reference models dedicated for the "Target Mobile Enterprise" phase are outlined later.

5.1 Mobile Vision Cube

The cube suggested is a representation of the three key parameters, when combined; we can identify mobility initiatives, mobile applications or mobile application families. The first parameter is the "User Need", the second is the enterprise "Value Chain" and the third is the "Device and Information System Capabilities".

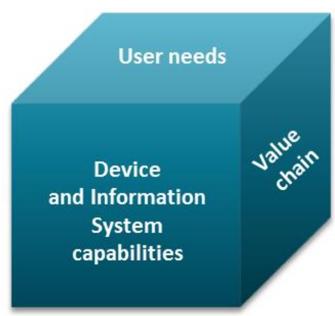


Fig. 2. Mobile Vision Cube

Mobility is, of course, introduced as an answer to one or many user needs. These users could be customers, employees or partners. This answer to user need is necessarily given within one of the enterprise value chain activities.
On the other side, mobile devices capabilities, the enterprise information system or Cloud capabilities could bring rich technology solutions to mobile use.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 1, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

To illustrate the use of the Mobile Vision Cube for an insurance company, an Insurance Claim mobile application is based on the combination of:

- A user need which is making an insurance claim quickly;
- The activity of "Insurance Claim Management" which is one of the most important insurance companies value chain activities.
- Capabilities of mobile devices like the feature of tacking photos with the mobile camera. The application should be, also, connected to the insurance company information system providing services of an insurance claim management system (IS capabilities);

Another example could be a mobile application for geo-localizing gas stations:

- The user need is locating the nearest gas station
- The value chain activity concerned within a gas distribution is "Customer Relationship Management"

And the capabilities that could be used are: mobile geo-localization capability (using GPS feature for example), gas station location and information delivered by the information system and a map retrieved from the Cloud.

The third example is Business Intelligence mobile applications family:

- The user (company managers in this case) need could be to have quick access to business key performance indicators;
- The value chain activity concerned could be the company "Sales Management" activity;
- The device capability is the possibility of handling business information through a local database updated from the enterprise information system.

5.2 Mobile value chain

The business value chain is a tool that helps identifying high level business processes or finer level (level 2 or 3) processes that are likely to be equipped with mobility initiatives and typical mobile applications. The activities of the Customer Relationship Management, for example, can be equipped with applications "Mobile CRM". Performance monitoring activities can be provided with "Mobile Business Intelligence" applications.

This mobile value chain can be developed for a given business (eg, commercial enterprises) or a given industry activity (insurance sector for example).

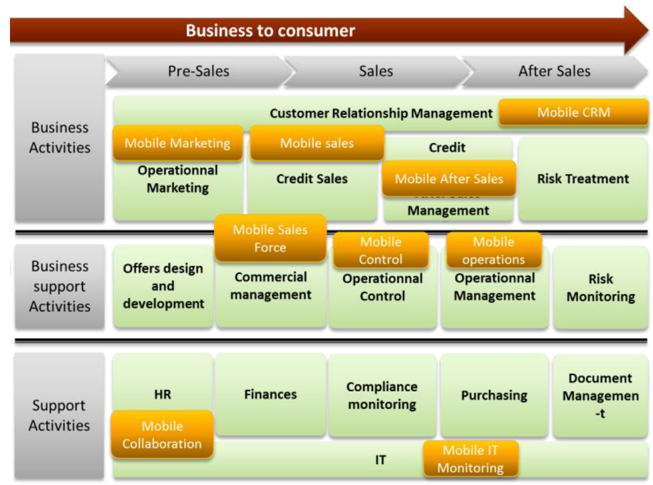


Fig. 3. Mobile value chain

5.3 Mobile Application reference model

The application reference model provides predefined mobile applications that can be implemented either at company's front-office domain or at the back-office one or in a in a transverse manner through the company.

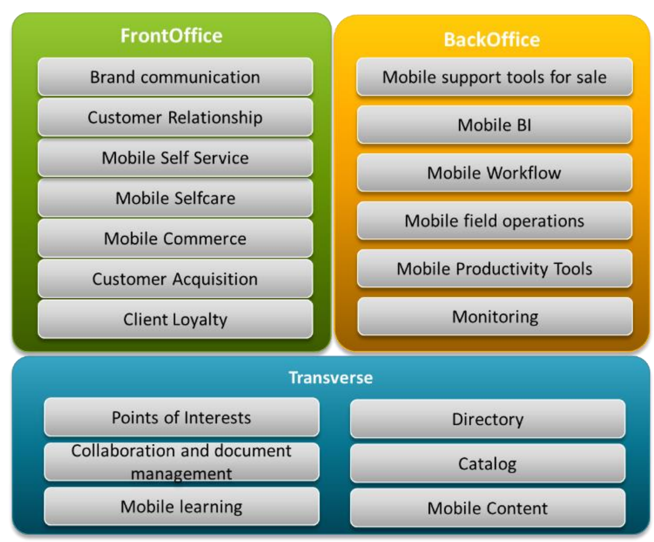


Fig. 4. Mobile application reference model

5.4 Mobile technical reference model

The mobile technical reference model lists a catalog of technological capabilities offered by mobile devices such as:

• The **sensors** (GPS, camera, accelerometer, etc.).

• **Standard applications** (Calendar, Contacts, Notification, etc.).

• **Entries** (touchscreen, keyboard input, voice recognition, etc.).
• Means of **communication** (voice call, sending SMS / MMS, sending email, etc.).
• **Outputs** (Text display, image and video display, sound, etc.).
• Means of **storage** (disk storage, memory storage, etc.).
• **Ports** (USB / Mini USB, TV output, Bluetooth, etc.).

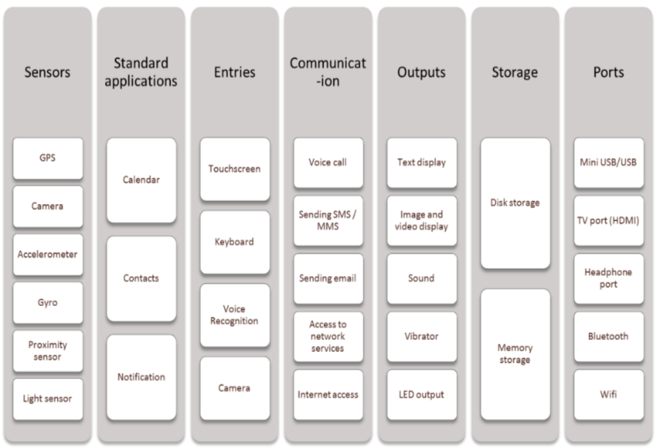


Fig. 5. **Mobile technical reference model**

6. Conclusion and future work

The methodology that has been proposed in this paper is inspired by the methodology TOGAF ADM while adding the specificities of mobile, lightweight appearance, pragmatic and progressive adapted to the requirements of business competitiveness. It is supported by three methodological tools and techniques to ensure a rapid and effective deployment of companies of different sizes and activities.
Methodological techniques (including models of application and technical references) that have been introduced are independently an extension and enhancement of existing architecture frameworks such as TOGAF to support the mobility aspects.
The methodology and the media are an important contribution in the mobile world that is often characterized by opportunism initiatives and time constraints imply a lack of organization and structure.
The methodology can be enriched by other methodological tools and techniques particularly for the phases Roadmap and Governance. It is also possible to define a maturity model for enterprise mobility that will define the content of VERGA cycle depending on the maturity of the company. It also plans to enhance the metamodel that we have defined in [17] to cover mobility aspect.

Références

[1] Chen G. and Kotz D. A Survey of Context-Aware Mobile Computing Research. Technical Report TR2000-381, Dartmouth Computer Science, November, 2000
[2] Ask J. The Future Of Mobile Experiences Is Context, Forrester Report, October 2011
[3] Schilit B. Adams N. Want R., Context-aware computing applications. In Proceedings of IEEE Workshop on Mobile Computing Systems and Applications, pages 85-90, Santa Cruz, California, December 1994. IEEE Computer Society Press.
[4] Chandler, Alfred A., Jr. (1962), Strategy and Structure: Chapters in the History ofAmerican Industrial Enterprise, The MIT Press, Cambridge, MA
[5] Porter, M. E. (1985). Competitive Advantage: Creating and Sustaining Superior Performance, Collier Macmillan, New York, N.Y.
[6] Pant S., Hsu C.: Strategic Information Systems Planning: A Review, 1995 Information Resources Management Association International Conference.
[7] The Open Group .: "TOGAF as an Enterprise Architecture Framework", http://www.opengroup.org/architecture/togaf8-doc/arch/
[8] Sessions R. : "A Comparison of the Top Four Enterprise-Architecture Methodologies", http://msdn.microsoft.com/en-us/library/bb466232.aspx (2007)
[9] Schekkerman J. : Trends *in Enterprise Architecture 2005: How are Organizations Progressing?*
[10] Arnoldo C. Hax (1996), "Strategy Concept and Process: A Pragmatic Approach", Barnes and Noble, bn.com
[11] Lederer, A. L. and Sethi, V. (1988). The Implementation of Information Systems Planning Methodologies, MIS Quarterly, September 1988, 445-461.
[12] Vitale, M., Ives, B. and Beath, C. (1986), "Identifying Strategic Information Systems," Proc. 7th Int'l Conf. Inf. Sys., San Diego, December 1986, pp. 265-276.
[13] CCTA (1999). IS Strategy: process and products, Format Publishing Limited, Norwich.
[14] Boar, B. (2001). The Art of Strategic Planning for Information Technology, John Wiley and Sons, New York, NY.
[15] Clevenger N., iPad in the Enterprise: Developing and Deploying Business Applications, John Wiley & Sons, 2011
[16] Lakhdissi M, Bounabat B (2011). "Toward a novel methodology for IT Strategic Planning". Proceeding of ICIME 2011, p 277-287, , Toronto Canada, ISBN: 97-1-906638-98-6, April 2011
[17] Lakhdissi M, Bounabat B (2011). "A new content metamodel for IS Strategic Planning with a tool implementation" International Jounal of Computer Science Issues(IJCSI) Volume 8 Issue 8, Mars 2012

Algorithms for Data Compression in Wireless Computing Systems

Upasana Mahajan, Dr. Prashanth C.S.R

CSE, VTU, Bangalore, Karnataka, India

Professor & HOD, Dept. of CSE, NHCE, VTU,
Bangalore, Karnataka, India

Abstract

Compression is a technique used for reducing data size by manipulating data redundancy; so that the packet transmission time and storage cost can be reduced. This can be achieved with the use of suitable data compression algorithms. Choosing the right algorithm can be accomplished by analyzing the performance of the algorithm. This paper presents the survey of various lossless data compression algorithms.

Keywords: *Lossless data compression, irreducible substitution tables, wireless sensor networks, compression algorithms, compression ratio.*

1. INTRODUCTION

Data Compression can be defined as encoding the information using the small number of bits rather than original representation. There are two types of data compression, lossless and lossy compression. The lossy compression is a method of data encoding, in which compression is done by discarding/losing some data. This is commonly used in multimedia data, especially in applications like streaming media and internet telephony. In this some loss of information is acceptable. Dropping nonessential detail from the data source can save storage space. There are two basic lossy compression schemes[1]lossy transform codecs and lossy predictive codecs. The lossless data compression can be defined as reducing the bits by identifying and eliminating statistical redundancy. The lossless data compression is reversible of lossy compression, such that the exact original data to be reconstructed from the compressed data. Lossless compression can be used for images, audios etc. but mostly it is used for text data like executable program, text documents and source code. In this paper focus is only on the lossless data compression. There are different types of lossless data compression algorithms[2] like Huffman's coding, Run Length encoding, Dictionary coders (LZW) etc. Based on the algorithm performance factors[10] like compression ratio, saving percentage and compression time, we choose the algorithm for compressing the data. The ultimate goal is to study different algorithms and select the best for compression.

2. LITERATURE SURVERY

2.1. Lossless Data Compression Algorithms Based on Substitution Tables [3] [4]

This paper introduces a class of new lossless data compression algorithm. Each algorithm first tries to transforms the original data, which is to be compressed into an irreducible table representation and then uses an arithmetic code to compress the irreducible table representation. These are generally known as universal coding algorithms as they try to achieve the compression rate. These new range of lossless data compression algorithm has been developed to improve overall compression rate and performance with the help of different variants of hierarchical transformations.

Firstly, the tables are formed with the help of parallel substitution which ends up with a unique string using the reduction rules. In this research 5 different reduction rules have been implemented and with the help of which, less complex tables are formed. For example, Let x be a string from A which is to be compressed. Starting from the table T consisting of only one row (s, x), a hierarchical transformation applies repeatedly the reduction rules 1-5 in some order to reduce T to an irreducible substitution table. To compress x, the corresponding algorithm then uses the zero order arithmetic code to compress the irreducible table. After receiving the code word of T', one can fully recover T' from which x can be obtained via parallel substitution. Some examples of hierarchical transformation are Greedy Sequential Transformation, SEQUITUR Transformation, Multilevel Pattern Matching Transformation (MPM),

The greedy sequential transformation parses the sequence, into non-overlapping substring and build a sequentially an irreducible table for each substring. This algorithm helps in sequential compression. The SEQUITUR algorithm has two main rules:

1. No pair of adjacent symbols appears more than once in the grammar.
2. Every rule is used more than once.

This helps to build irreducible table for each prefix and then append a substring to the end of the row at last apply the reduction rules 1-5 to reduce the table. It transforms the binary sequence.

The MPM transformation bisects each distinct substring repeatedly, until the length of substring is 2. Then, assign a unique token to each substring and create a substitution table. The MPM code and Lempel-Ziv code have similarities like both are pure pattern matching codes, so they do not directly compress the data. But there are differences like MPM is a hierarchical transformation so it does pattern matching at multiple levels and the LZE is non-hierarchical. The MPM code was developed for, strictly for data of length a power of two, and named the bisection algorithm.

This research helps in trying to solve a problem of performance of an algorithm. It can be evaluated mainly by calculating and comparing the two facts: Frequency of a block of a sequence and Empirical Entropy of a sequence.

2.2. A Simple Algorithm for Data Compression in Wireless Sensor Networks [5] [6]

Sensor Nodes have small batteries which cannot be changed or recharged frequently, so the WSN have an issue of Energy. Power saving can be done by either duty cycling (coordinated sleep/wakeup schedules between nodes) or by in-network processing (compression/aggregation techniques). Data compression is the best option and appreciated only if the execution of compression algorithms requires lesser amount of energy than the one saved in reducing transmission. This paper introduces the algorithm known as Lossless Entropy Compression (LEC), which shows the correlation between the data collected by sensor nodes and the entropy compression. This algorithm follows same scheme used in baseline JPEG algorithm for compressing the DC-coefficients of a digital image. The Huffman table proposed in JPEG to entropy encoding the groups has been adopted.

```
Encode (d_i, Table)

 IF d_i=0 THEN
    SET n_i = 0
 ELSE
    SET n_i = ⌈ log2(|d_i|) ⌉
ENDIF
    SET s_i TO Table [n_i]
IF n_i= 0 THEN
    SET bs_i = s_i
ELSE
  IF d_i > 0 THEN
    SET a_i= (d_i)|ni
ELSE
     SET a_i = (d_i – 1)|ni
   ENDIF
  SET bs_i TO <<s_i, a_i>>
  ENDIF
RETURN bs_i
```

Fig.1 Pseudo-code of the encode algorithm

The difference d_i computed by algorithm for the input to an entropy encoder. The $d_i = bs_i$ (bit sequence) = $s_i|a_i$, s_i codifies the number n_i and a_i represents d_i. If

i. $d_i > 0$, $a_i = n_i$ low order bits of the 2's complement representation of d_i.
ii. $d_i < 0$, $a_i = n_i$ low order bits of the 2's complement representation of $d_i – 1$.
iii. $d_i = 0$, s_i is coded as 00 and a_i is not represented.

Table1: Huffman variable length codes used in the experiment.

n_i	s_i	d_i
0	00	0
1	010	-1,+1
2	011	-3,-2,+2,+3
3	100	-7..-4,+4..+7
4	101	-15..-8,+8..+15
5	110	-31..-16,+16..+31
6	1110	-63..-32,+32..+63
7	11110	-127..-64,+64..+127
8	111110	-255..-128,+128..+255
9	1111110	-511..-256,+256..+511
10	11111110	-1023..-512,+512..+1023
11	111111110	-2047..-1024,+1024..+2047
12	1111111110	-4095..-2048,+2048..+4095
13	11111111110	-8191..-4096,+4096..+8191
14	111111111110	-16383..-8192,+8192..+16383

EXPERIMENTAL RESULTS:

The performance of a compressed algorithm can be defined by compression ratio as shown below:

$$\text{Comp\_Ratio} = 100 \text{ X } 1 - \left[\frac{\text{Comp\_Size}}{\text{Orig\_Size}}\right]$$

	Temperature		Relative Humidity	
	Comp_size	*Comp_ratio*	*Comp_size*	*Comp_ratio*
LEC	7605bits	66.99%	7527bits	67.33%
S-LZW	16760bits	27.25%	13232bits	42.57%
Gzip	15960bits	30.73%	13320bits	42.19%
Bzip2	15992bits	30.59%	13120bits	43.05%

With the help of datasheets given in SHT11 for temperature & relative humidity and using the above formula, following ratios are obtained. Thus the comparison between other compression algorithm ratios as per following results shows that the LEC algorithm performs better.

2.3. Online Adaptive Compression in Delay Sensitive Wireless Sensor Networks [7] [8]

In wireless sensor networks (WSN), compression reduces the data size by exploiting the redundancy residing in sensing data. This reduction of the data can be measured as compression ratio which is calculated as original data size divided by the compressed data size. The higher the compression ratio means more data reduction is done and results in shorter communication delays. To understand the effect of compression, firstly obtain the processing time of compression, which depends on several factors like compression algorithm, CPU frequency, processor architecture and compression data. There are so many compression algorithm have been developed, but one of the best is Lempel-Ziv-Welch (LZW). LZW is a dictionary based lossless compression algorithm suitable for sensor nodes which replaces the strings of characters with single codes in the dictionary. To calculate the compression delay, the software estimation approach is adopted. The source code of this algorithm is written in C and then converted into the assembly codes, which have fixed number of execution cycles.

```
STRING = get the first character

while there are still input character

 C = get next character

look up STRING + C in the dictionary
if STRING+C is in the dictionary
        STRING = STRING + C
else
        output the code for STRING
        add STRING+C to the dictionary
        STRING = C
        end if
end while
output the code for STRING
```

Fig. 2 the LZW Compression Algorithm

The total count of cycles can be obtained at the completion of LZW algorithm. The processing time of the algorithm can be calculated by dividing the total execution cycles by the working (i.e. CPU) frequency. There are different experiments conducted in the NS-2 simulator to check out the effect of compression on the packet delays. The results of the experiments shows that delay can cause severe performance degradation under light traffic load and if traffic loads is heavy than compression reduces the delay of packet, increase the maximum throughput. So the compression is favored only when the packet generation rate is higher than the threshold rate. Therefore to determine whether the compression of data is required or not the online adaptive algorithm has been developed.

The adaptive compression algorithm is distributively implemented on each sensor node as ACS (Adaptive Compression Service) in an individual layer created in a network stack. The main goal of this algorithm is to take a right decision, that whether packet transmission is required or not at a particular node. Before moving to algorithm, let's have a look of the architecture of ACS. There are 4 functional units: 1) **Controller** manages the traffic flow and makes compression decisions on each incoming packet in this layer. 2) The **LZW compressor** performs actual compression of packet with the help of LZW algorithm. 3) The **information collector** helps in collecting local statics information about network and hardware conditions. 4) The **packet buffer** helps in temporarily storing the packets to be compressed.

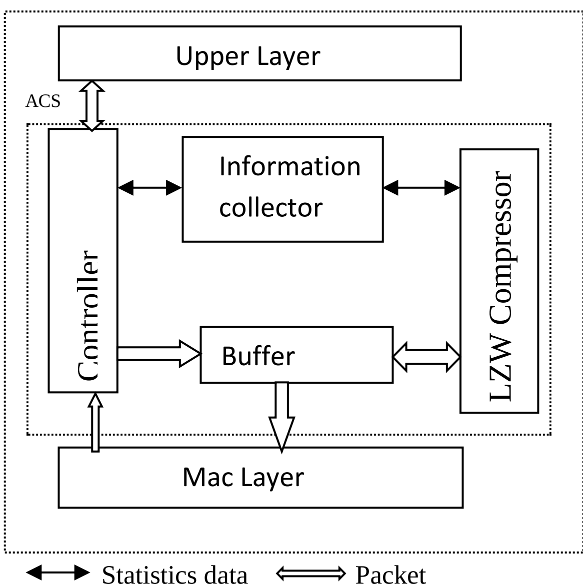


Fig.3 Architecture of ACS

As compression is managed by the node state, so the adaptive algorithm helps to determine the node state. In this algorithm the utilization of the queuing model is done for estimation of the node state conditions. The queuing model includes the network model and the MAC model. The network model defines the network topology and traffic (i.e. estimates the arrival rates of each node). The MAC model defines the packet service time with the help of DCF (Distributed Coordination Function), which can be calculated as the time when packet enters the MAC layer to the time when packet is successfully transmitted or discarded.

The Adaptive Compression Algorithm is divided into two stages: Information collection and State determination. Firstly, in ACS the information collector collects the statistics information like compression statistics (compression ratio r_c, average compression processing time T_p, the coefficient of variance of processing time c_p), MAC layer service time and packet arrival rates. Once the collector finishes its job, the controller in the ACS defines the state of the node i.e. whether compression is required or not. For making the decision the following State Determination Procedure has been adopted which is performed at the end of each time slot for a node in a No-Compression state.

For each node at level i:
if state = No-Compression **then**
read statistics from the information collector
compute T_{com}, ΔT_{min}
if $T_{com} \leq \Delta T_{min}$ **then**
set state to Compression
else
    set i to the node's level number
    $\Delta T_{mac} = 0$
    while i > 0
    calculate λ^i and
compute reduction ΔT_{mac} (i)
    add λ^i ΔT_{mac} (i) to ΔT_{mac}
    decrease i by one
    end while
    if $\lambda_c T_{com} \leq \Delta T_{mac}$ then
    set state to Compression
    end if
end if

Fig. 4 State Determination Procedure

T_{com} the average packet waiting time at the compression queue.
ΔT_{min} lower bound of total delay reduction
ΔT_{mac} MAC layer service time
$\Delta T_{mac}(i)$ Delay reduction in level i
λ_c Arrival rate compression
λ^i Mean arrival rate for nodes in level i

With help of the queuing model, it is possible to calculate the terms/equation used in algorithm. So the outcome this paper is that using the online adaptive compression algorithm, each node can decide whether the packet is compressed or not, adapting to the current network and hardware environment.

2.4. A Statistical Lempel-Ziv Compression Algorithm for Personal Digital Assistant (PDA) [9]

This paper introduces a compression algorithm named as Statistical Lempel-Ziv Compression algorithm (SLZ), which is suitable for the applications of hand held PDAs and can be viewed as a variant of LZ77.

The first step of the algorithm is to build a dictionary which may include up to $2^{|c|}$ entries (supposing each fixed length codeword *c* contains |c| bits). To build a good dictionary, a two pass approach is adopted. The first pass is to collect most useful phrases from the file for building a dictionary. The second pass is to do compression by creating codewords that refer to the phrases in the dictionary. While building a dictionary, there must be a balance between the dictionary size and codeword length to avoid large number of phrases. Therefore for a file of *T* symbols long, the total number of phrases will be:

$$\sum_{i=0}^{T-1} T\text{-}i = \frac{T^2 + T}{2} \approx O(T^2) \qquad (1)$$

The sliding window approach which has been used in LZ77 can be adopted to reduce the number of phrases. Let's imagine the sliding window of size W symbols, such that W<<T then

$$\sum_{i=0}^{W-1} T\text{-}i = \frac{W(2T - W + 1)}{2} \approx O(WT) \qquad (2)$$

Once done with the number of phrases, time is to decide which phrases have to collect. If the phrase collected from the file is found in the dictionary, then there is no need to add that phrase in the dictionary but the number of counts/ frequency of that phrase is incremented. On the other hand, two identical phrases having overlap in the input file must be counted as single occurrence instead of two. This overlap detection can be done by adding a time stamp (last time at which the phrase occurred in the file) to each entry in the dictionary. When a phrase is fetched from the input file and an identical phrase is found in the dictionary, compare the timestamp of that phrase in the dictionary with the current time stamp. If time stamp difference is less than the phrase length, overlap is detected.

After collecting the phrases, time to put all phrases in the dictionary with respect of dictionary size. The size of dictionary should not be too large and too small; it must contain all useful phrases. The number of entries in the dictionary can be reduced by pruning the phrases having unit frequency. Which means prune the phrases which occur once in a file not the phrases that are one symbol long, and it can be done at end of the first pass. With this method most of the time the newly appears phrases are purged. So to avoid this problem another method of pruning the phrases known as Move-To-Front approach is used.

In this approach, when a new phrase inserted it is move to the front of the dictionary. The time dictionary is full, discard the phrase at the end of the dictionary. With this method the phrases which have high frequency will be at the front and phrases with least frequency located at the end of the dictionary. Once the dictionary has been build, it's time for compression which can be done with the help of entropy coding method. The symbols of the file are shifted into the sliding window and once it's full, the symbol sequence in window is compared with the phrases one by one in the dictionary in the order of entropy. When matched phrase is found, the matched symbols in the window are coded by the index of that phrase. The symbols that matched the phrases are removed from the sliding window and new symbols are moved to sliding window. As soon as window is full repeat the process again until all the symbols get coded.

We can't say that this is the best compression algorithm but a simple entropy coding scheme designed using the prefix codes to eliminate look-up table for decoding. Using the combination of dictionary based algorithm and sliding window approach, the overall compression ratio decreases.

2.5. Comparison of Lossless Data Compression Algorithms for Text Data [10]

Data compression helps in reducing the size of the file, in other words compression represents the information in a compact form rather than its original form without any data loss. When data compression is done while transmitting the data, the main concern is speed. Speed of the transmission depends upon the number of bits sent, the time required for the encoder to generate the coded message and the time required for the decoder to recover the original message. Sometimes the lossless compression algorithms are also known as reversible algorithms, as the original message can achieve by the decompression process. Some of the famous lossless compression algorithms are Run-Length Encoding (RLE), Huffman Encoding, Adaptive Huffman Encoding, Shannon Fano algorithm, Arithmetic Encoding and Lempel Zev Welch algorithms.

This paper introduces the comparison of performances of above algorithms, based on different factors. There are many different ways to measure the performance of a compression algorithm. The main concern is space and time efficiency, while measuring the performance. Following are some factors used to evaluate the performances of the lossless algorithms.

$$\textbf{Compression Ratio} = \frac{\textbf{size after compression}}{\textbf{size before compression}} \qquad \textbf{(3)}$$

$$\textbf{Compression Factor} = \frac{\textbf{size before compression}}{\textbf{size after compression}} \qquad \textbf{(4)}$$

$$\textbf{Saving Percentage} = \frac{\textbf{size before compression} - \textbf{size after compression}}{\textbf{size before compression}} \% \qquad \textbf{(5)}$$

Compression Time can be defines as time taken to compress particular file. Time taken for the compression and decompression should be considered separately. For a particular file, if the compression and decompression time is less and in an acceptable level, it means that algorithm is acceptable with respect to time.

Entropy can be used as a performance factor, if the compression algorithm is based on statistical information of the source file. Let set of event be S= $\{s_1, s_2, s_3 \ldots s_n\}$ for an alphabet and each s_j is a symbol used in this alphabet. Let the occurrence probability of each event be p_j for event s_j. Then the self-information I(s) is defined as follows:

$$I(s) = \log_b 1/p_j \text{ or } I(s) = -\log_b 1/p_j \qquad \textbf{(6)}$$

The first order Entropy value H(P) can be calculated as follows:

$$H(P) = \sum_{j=1}^{n} p_j I(s_j) \quad \text{or} \quad H(P) = -\sum_{j=1}^{n} p_j I(s_j) \qquad \textbf{(7)}$$

Code Efficiency is the ratio between the entropy of the source and the average code length.

$$E(P, L) = \frac{H(P)}{\bar{l}(P, L)} 100\%, \qquad \textbf{(8)}$$

E (P, L) is the code efficiency, H (P) is entropy and $\bar{l}(P, L)$ is the average code length.

Average code length defined as the average number of bits required to represent a single code word. It can be calculated as: $\bar{l} = \sum_{j=1}^{n} p_j * l_j$, where p_j is the occurrence probability of j^{th} symbol of the source message, l_j is the length of the particular code word for that symbol and $L = \{l_1, l_2, \ldots\ldots, l_n\}$.

In order to test the performance of above mentioned lossless compression algorithms, first step is to implement them and then test them with some set of files. Performances evaluated by computing above mentioned factors. After the implementation and testing the results shows that the Adaptive Huffman algorithm needs larger time period for processing, because the tree should be updated or recreated. LZW works better as the file size grows up to certain limit, because there are more chances of replacing the words by using the small index number. But it cannot be used for all cases, so can't say it is one of the efficient algorithms.

Arithmetic Encoding algorithm has an Underflow problem, which gives an erroneous result after few numbers of iterations. Therefore it is not suitable for comparison. Huffman Encoding and Shannon Fano algorithm shows similar results except in compression times. Shannon Fano algorithm has faster compression time than Huffman Encoding, so this factor can be used to determine the more efficient algorithm from these two.

While considering the major performance factors like compression time, decompression time and saving percentages of the all the selected algorithms. The Shannon Fano algorithm is considered as the most efficient algorithm, as the values of this algorithm lies acceptable range and it also shows better results for the large files.

3. CONCLUSION

This study introduces data compression and simple algorithms for compression. Each algorithm has its own advantages and disadvantages. With the help of various performance factors, it is easy to choose algorithms that are more efficient. This paper demonstrates that if we use the right data compression techniques, it will certainly be helpful in reducing the storage space and the computational resources. This is definitely more critical in the case of wireless systems where network bandwidth is always a cause for concern.

REFERENCES

[1] https://en.wikipedia.org/wiki/Lossy_compression.
[2] https://en.wikipedia.org/wiki/Lossless_ Compression.
[3] John.C.Kieffer and En-hui Yang, "Lossless Data Compression Algorithms Based on Substitution Tables", IEEE, 1998.
[4] J.C.Kieffer, E-H.Yang, G.Nelson, and P.Cosman, "Lossless data compression via multi-level pattern matching," IEEE, 1996.
[5] Massimo Vecciho, "A Simple Algorithm for Data Compression in Wireless Sensor Networks", IEEE, June 2008.
[6] https://en.wikipedia.org/wiki/Huffman_coding.
[7] Xi Deng and Yuanyuan Yang, "Online Adaptive Compression in Delay Sensitive Wireless Sensor Networks", IEEE, October 2012.
[8] https://en.wikipedia.org/wiki/Lempel–Ziv–Welch.
[9] S.Kwong and Y.F.Ho, "A Statistical Lempel-Ziv Compression Algorithm for Personal Digital Assistant (PDA)", IEEE, February 2001.
[10] S.R.Kodituwakku and U.S.Amarasinghe, "Comparison of Lossless Data Compression Algorithms for Text Data", Indian Journal of Computer Science and Engineering, Vol 1 No. 4 416- 425.

Upasana Mahajan is M.Tech in Computer Science Engineering (2006) from VTU and passed B.Tech (CSE) from Guru Nanak Dev University in year of 2006. She is currently pursuing her Ph.D and has active research interest in area of Data Compression and there applications in wireless networks.

Dr. Prashanth C.S.R holds B.E, M.S (USA), Ph.D (USA). At present he is a Professor and Head of the Department for Computer Science in New Horizon College of Engineering Bangalore, Karnataka.

A Generalized symmetric single step method for solving interval linear systems.

Hassan Badry Mohamed A., El-Owny

Mathematics department, Faculty of Science, Aswan University, 81528 Aswan, Egypt.
CIT College, Taif University, 21974 Taif, KSA.

Abstract

Systems of algebraic equations with interval coefficients are very common in several areas of engineering sciences. Generalized intervals extend classical intervals providing better algebraic properties. These properties allow one constructing a generalized symmetric single step method. This paper proposes a new C-XSC (C- for eXtended Scientific Computing) software for the symmetric single step method with generalized intervals for computing an enclosure for the solution set. Examples illustrating the applicability of the proposed method are solved, and compared with other methods.

Keywords: *interval linear systems, validated interval software, C-XSC, symmetric single step method.*

1. Introduction

Solving linear systems involving uncertainties in the parameters is an important part of the solution to many scientific and engineering problems. But in real life situations, parameters of these systems often are charged by different kinds of uncertainties[4, 5, 6]. Leontief's input output model of economy[16] can be taken as an example[29]. In many cases, uncertainty can be represented by intervals. Since the seminal publication by Moore's[17], a rapid development of interval arithmetic had been observed. The system of linear interval equations can be presented as follows:

$$[A]\cdot x=[b], \tag{1}$$

with $x \in \mathbb{R}^n$, the matrix $[A] \in \mathbb{IR}^{n\times n}$ and the vector $[b] \in \mathbb{IR}^n$ are said to belong to interval family if their elements are from some real intervals. System (1) is called the interval system of equations. Suppose $[A]$ is regular i.e. A is nonsingular for any $A \in [A]$. Then for a matrix $A \in [A]$ and any vector $b \in [b]$ an ordinary linear system $A \cdot x = b$ has the unique solution. We are interested in a set $\sum$ of all these solutions of interval system:

$$\sum = \{x \in \mathbb{R}^n : Ax = b, A \in [A], \quad b \in [b]\}$$

The characterization of this set has been obtained in[20]. It has been proved that the intersection of $\sum$ with each orthant in $\mathfrak{R}^n$ gives a convex polytope. But in general, $\sum$ is non-convex as the union of convex sets, and its detailed description meets combinatorial difficulties. The main objective is to find interval solution of linear interval system that is to determine the smallest interval vector [y] containing all possible solutions. In other words, we need to imbed the solution set $\sum$ into the minimal box in $\mathbb{R}^n$. This problem is known to be NP-hard [15] and complicated from computational viewpoint for large-scale systems. Oettli[21] shows how multiple linear programming can be used to obtain [y]; this line of research was continued by Cope and Rust[7], and Rust and Burrus[26]. Some iterative approaches were established t this context as well as direct numerical methods that provide over-bounding of [y] (see monographs[8, 9, 19] and papers[25, 27]).

In this paper, we propose a new C-XSC software (C for Extended Scientific Computing)[10] of the symmetric single step method for the solution of the systems of interval linear equations with the use of „generalized interval". We will compare our method to other methods. The rest of the paper is set as follows. In Section 2, some Basic notation is introduced. In section 3, we recall the fundamentals of „generalized interval" and present its interpretation as the modified of interval arithmetic. The main results of this paper is presented in Section 4. Another modification for the symmetric single step method is introduced in Section 5. Numerical and practical examples illustrating the features of the proposed method are provided in Section 6. Last section concludes with some remarks.

2. Basic notations

We use the following notations $\mathbb{R}, \mathbb{R}^n, \mathbb{R}^{n\times n}$, $\mathbb{IR}, \mathbb{IR}^n, \mathbb{IR}^{n\times n}$, to denote the set of real numbers, the set of real vectors with n components, the set of real $n\times n$ matrices, the set of intervals, the set of interval vectors with n components and the set of $n\times n$ interval matrices, respectively. By interval we mean a real compact interval

$$[x] \coloneqq [a,b] \coloneqq \{x \in \mathbb{R} | \quad a \le x \le b\}$$

For $[x], [y] \coloneqq [c, d]$ we define

- The mid-point $\mathrm{mid}([x]) \coloneqq {(a+b)}/{2}$,
- the radius $\mathrm{rad}([x]) \coloneqq {(b-a)}/{2}$,
- the absolute value $|[x]| \coloneqq \max\{|a|, |b|\}$, ,
- the distance $q([x],[y]) \coloneqq \max\{|a-c|, |b-d|\}$

- minimal absolute value (mignitude)

$$\prec [x] \succ := \min\{|x| : x \in [x]\} = \begin{cases} \min\{|a|,|b|\} & \text{if } 0 \notin [x] \\ 0 & \text{else} \end{cases} \quad (2)$$

For interval vectors and interval matrices, these quantities are defined componentwise. If for two interval vectors $[u],[v] \in \mathbb{IR}^n$ we have $[u_i] \cap [v_i] \neq \emptyset, i = 1,2,\cdots,n,$ then $[u] \cap [v] := ([u_i] \cap [v_i])$ otherwise $[u] \cap [v] = \emptyset$. In addition, for $[u],[v] \in \mathbb{IR}^n$ we define $[u] \subseteq [v$ iff $[u_i] \subseteq [v_i], i = 1,2 \cdots, n$. Furthermore, we repeat some relations concerning the distance:

$$q([u],[v]) \leq q([u],[w]) + q([w],[v])$$
$$q([u]+[w],[v]+[w]) = q([u],[v])$$
$$q([u]+[v],[w]+[z]) \leq q([u],[w]) + q([v],[z])$$

if $[u],[u],[w],[z] \in \mathbb{IR}^n$.

For square interval matrices we define the comparison matrix (Ostrowsky matrix) $\langle [A] \rangle := (C_{ij}) \in \mathbb{R}^{n\times n}$ using (2) by setting

$$c_{ij} := \begin{cases} -|a_{ij}| & \text{if} \quad i \neq j \\ \langle [a_{ij}] \rangle & \text{if} \quad i = j \end{cases}$$

A square matrix $[A] \in \mathbb{IR}^{n\times n}$ is called regular if all $A \in [A]$ are nonsingular.

If $\mathrm{mid}([A]) \cdot [A]$ is regular then $[A]$ is strongly regular. An interval matrix $[A]$ is an H-matrix iff there exist a vector $v > 0$ such that $\langle [A] \rangle v > 0$.

Definition 1 [28] *Let* $A,B,C \in \mathbb{R}^{n\times n}$. *Then* $A = B - C$ *is a regular splitting of* A *if* $C \geq 0$ *and* B *is nonsingular with* $B^{-1} \geq 0$.

Theorem 1 [28] *Assume that is* $A \in \mathbb{R}^{n\times n}$ *nonsingular, that* $A^{-1} \geq 0$ *and that* $A = B - C$ *is a regular splitting of* A. *Then* $\rho(B^{-1}C) < 1$, *where* $\rho(\ )$ *denotes the spectral radius of a matrix.*

Regular splitting was introduced in [28], where one can also find the proof of Theorem 1.

3. Generalized intervals

Generalized intervals are intervals whose bounds are not constrained to be ordered, for example [-2,2] and [2,-2] are generalized intervals. They have been introduced in[12,22] so as to improve the algebraic structure of intervals, while maintaining the inclusion monotonicity. The set of generalized intervals is denoted by $\mathbb{KR}$ and is divided into three subset:

- The set of *proper intervals* with bounds ordered increasingly. These proper intervals are identified with classical intervals. The set of proper intervals is denoted $\mathbb{IR} := \{[a,b] | a \leq b\}$. *Strictly* proper intervals satisfy $a < b$.
- The set of *improper intervals* with bounds ordered decreasingly. It is denoted by $\overline{\mathbb{IR}} := \{[a,b] | a \geq b\}$. *Strictly* improper intervals satisfy $a > b$.
- The set of *degenerated intervals* $\{[a,b] | a = b\} = \mathbb{IR} \cap \overline{\mathbb{IR}}$. Degenerated intervals are identified to reals.

Therefore, form a set of reals $\{x \in \mathbb{R}^n | a \leq x \leq b\}$, one can build the two generalized intervals [a,b] and [b,a]. It will be convenient to switch from one to the other keeping the underlying set of reals unchanged. To this purpose, the following three operations are introduced:

- The dual operation is defined by $\mathrm{dual}([a,b]) = [b,a]$.
- The proper projection is defined by $\mathrm{pro}([a,b]) = [\min\{a,b\}, \max\{a,b\}]$.
- The improper projection is defined by $\mathrm{imp}([a,b]) = [\max\{a,b\}, \min\{a,b\}]$.

The generalized intervals are partially ordered by an inclusion which extends the inclusion of classical intervals. Given two generalized intervals $[x] = [\underline{x}, \overline{x}]$ and $[y] = [\underline{y}, \overline{y}]$, the inclusion is defined by $[x] \subseteq [y] \Leftrightarrow \underline{y} \leq \underline{x} \wedge \overline{x} \leq \overline{y}$. For example, $[-1,1] \subseteq [-1.1,1.1]$ (this matches the set inclusion), $[-1.1,1.1] \subseteq [-1,1]$ (the inclusion between the underlying set of real is reversed for improper intervals) and $[2,0.9] \subseteq [-1,1]$. As degenerated intervals are identified to reals, if $[x]$ proper then $x \in [x] \Leftrightarrow x \subseteq [x]$. On the other hand, if $[x]$ is strictly improper then for all $x \in \mathbb{R}$ the inclusion $x \subseteq [x]$ is false.

The generalized interval arithmetic (Kaucher arithmetic) extends the classical interval arithmetic. Its definition can be found in[13,27]. When only proper intervals are involved, this arithmetic coincides with the interval arithmetic: $[x] \circ [y] = \{x \circ y \in \mathbb{R} | x \in [x], y \in [y]\}$. When proper and improper intervals are involved, some new expressions are used . For example, $[a,b] + [c,d] = [a+c, b+d]$ and if $a,b,c,d \geq 0$ then $[a,b] \cdot [c,d] = [a \cdot c, b \cdot d]$. The following useful property provides some bounds on the proper projection of the results of the generalized interval arithmetic. Let us consider $[x],[y] \in \mathbb{KR}$ and $\circ \in \{+,-,\cdot,/\}$. If $\mathrm{pro}[x] \circ \mathrm{pro}[y]$ is defined then $[x] \circ [y]$ is defined and it satisfies

$$\mathrm{pro}([x] \circ [y]) \subseteq (\mathrm{pro}[x]) \circ (\mathrm{pro}[y])$$

Generalized interval arithmetic has better algebraic properties than the classical interval arithmetic. The addition in $\mathbb{KR}$ is a group. The opposite of an interval $[x]$ is $-\mathrm{dual}[x]$, i.e.

$$[x] + (-\mathrm{dual}[x]) = [x] - \mathrm{dual}[x] = [0,0].$$

The multiplication in $\mathbb{KR}$ restricted to generalized intervals whose proper projection does not contain 0 is also a group. The inverse of such a generalized interval $[x]$ is $1/\mathrm{dual}[x]$, i.e.,

$$[x] \cdot (1/\mathrm{dual}[x]) = [x]/(\mathrm{dual}[x]) = [1,1] .$$

Although addition and multiplication in $\mathbb{KR}$ are associative, they are not distributive. The addition and multiplication in $\mathbb{KR}$ are linked by the following distributivity laws[24,27]. Whatever are $[x],[y],[z] \in \mathbb{KR}$

- Conditional distributivity:
  $[x]\cdot[y] + (\mathrm{imp}[x])\cdot[z] \subseteq [x]\cdot([y]+[z]) \subseteq [x]\cdot[y] + (\mathrm{pro}[x])\cdot[z]$.

The three following particular cases will be of practical interest in this paper.

- Subdistributivity: if $[x] \in \mathbb{KR}$ then $[x]\cdot([y]+[z]) \subseteq [x]\cdot[y]+[x]\cdot[z]$;
- Superdistributivity: if $[x] \in \overline{\mathbb{KR}}$ then $[x]\cdot([y]+[z]) \supseteq [x]\cdot[y]+[x]\cdot[z]$;
- Distributivity: if $x \in \mathbb{R}$ then $x\cdot([y]+[z]) = x\cdot[y]+x\cdot[z]$.

Another useful property of the Kaucher arithmetic is its monotonicity with respect to the inclusion. Whatever are $\circ \in \{+,-,\cdot,/\}$ and $[x],[y],[xx],[yy] \in \mathbb{KR}$,
$[x] \subseteq [xx] \cap [y] \subseteq [yy] \Rightarrow ([x]\circ[y]) \subseteq ([xx]\circ[yy])$.
The next example illustrates the way these properties will be used in the sequel.

Example 1: *Consider the expression* $[x]+[u][v] \subseteq [y]$. *Subtracting* $\mathrm{dual}([u][v]) = (\mathrm{dual}[u])(\mathrm{dual}[v])$ *to each side preserves the inclusion:* $[x]+[u][v]-\mathrm{dual}([u][v]) \subseteq [y]-(\mathrm{dual}[u])(\mathrm{dual}[v])$. *As* $-\mathrm{dual}([u][v])$ *is the opposite of* $[u][v]$, *the following inclusion is eventually proved to hold:* $[x] \subseteq [y]-(\mathrm{dual}[u])(\mathrm{dual}[v])$.

Finally, generalized interval vectors $[x] \in \mathbb{KR}^n$ and generalized interval matrices $[A] \in \mathbb{KR}^{n\times n}$ together with their additions and multiplications are defined similarly to their real and classical interval counterparts.

4. Generalized symmetric single step method

In this section we assume that the reader is familiar with the concept P contractions for proving the convergence of a fixed point iteration to a unique fixed point for an arbitrary starting vector. For the details please see [1,2].

We assume throughout that the matrix $[A]$ is nonsingular, and moreover that its diagonal entries $[a_{ii}]$ are all nonzero. We can express the matrix $[A]$ as the matrix sum [1,28]

$$[A]=[D]+[L]+[U],$$

Where $[D]$ is a diagonal matrix, and $[L]$ and $[U]$ are respectively strictly lower and upper triangular matrices. We can write (1) as

$$[D]\cdot x = [b] - ([L]+[U])\cdot x$$

Then

$$x=[D]^{-1}\cdot([b]-([L]+[U]\cdot x), \tag{3}$$

where

$$[L]=\begin{pmatrix} 0 & 0 & \cdots & 0 \\ [a_{21}] & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ [a_{n1}] & \cdots & [a_{nn-1}] & 0 \end{pmatrix} \tag{4}$$

$$[U]=\begin{pmatrix} 0 & [a_{12}] & \cdots & [a_{1n}] \\ \vdots & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & [a_{n-1n}] \\ 0 & 0 & \cdots & 0 \end{pmatrix} \tag{5}$$

$$[D]^{-1}=\begin{pmatrix} \frac{1}{\mathrm{dual}[a_{11}]} & 0 & \cdots & 0 \\ 0 & \frac{1}{\mathrm{dual}[a_{22}]} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\mathrm{dual}[a_{nn}]} \end{pmatrix} \tag{6}$$

Since the diagonal entries $[a_{ii}]$ of $[A]$ are nonzero, we can carry out the following iterative method derived from (3):

$$\left[x_i^{(l+1/2)}\right] = \frac{\left([b_i]-\sum_{j=1}^{i-1}\mathrm{dual}\left([a_{ij}]\left[x_i^{(l+1/2)}\right]\right)-\sum_{j=i+1}^{n}\mathrm{dual}\left([a_{ij}]\left[x_i^{(l)}\right]\right)\right)}{\mathrm{dual}([a_{ii}])}, \quad 1 \le i \le n \tag{7}$$

$$\left[x_i^{(l+1)}\right] = \frac{\left([b_i]-\sum_{j=1}^{i-1}\mathrm{dual}\left([a_{ij}]\left[x_i^{(l+1/2)}\right]\right)-\sum_{j=i+1}^{n}\mathrm{dual}\left([a_{ij}]\left[x_i^{(l+1)}\right]\right)\right)}{\mathrm{dual}([a_{ii}])}, \quad 1 \le i \le n,\ l \ge 0 \tag{8}$$

where the $[x^{(0)}]'s$ initial interval vector. We call this iteration procedure the **generalized symmetric single step method**.

Theorem 2. *Consider interval linear system (1), we define* $[L]$, $[U]$ *and* $[D]^{-1}$ *as in (4), (5) and (6), respectively. Then, the sequence* $\left[x^{(l)}\right]_{l=0}^{\infty}$ *calculated according to the iteration method (generalized symmetric single step method defined as in (7) and (8)), converges for all interval vectors* $[x^{(0)}] \in \mathbb{IR}^n$ *to* $[x^*]$, *where* $[x^*]$ *is the unique fixed point of the equation (3).*

5. Generalized symmetric single step method with intersection

In this section we consider modifications of the preceding iterative methods which are based on the fact that if for any of these methods one is starting with an interval vector containing the limit, then all iterates contain the limit. Therefore the enclosure of the limit might be improved by forming intersections after each iteration step.

Theorem 3. *Let* $[A] \in \mathbb{IR}^{n\times n}$ and $[b] \in \mathbb{IR}^{n}$ *be given.* $[L]$, $[U]$ *and* $[D]^{-1}$ *as in (4), (5) and (6), respectively. Let* $[x^*]$ *is the unique fixed point of the equation (3). We assume that we have an interval vector* $[\text{initial}] \in \mathbb{IR}^{n}$ *satisfying* $[x^*] \subseteq [\text{initial}]$. *We consider the generalized symmetric single step method with intersection.*

$$\left\{\begin{array}{l} [x^{(0)}] := [\text{initial}] \\ \textbf{for } i=1 \textbf{ to } n \textbf{ do} \\ \left[x_i^{(l+1/2)}\right] := [x_i^{(l)}] \cap \left\{ \left([b_i] - \sum_{j=1}^{i-1} \text{dual}\left([a_{ij}]\left[x_i^{(l+1/2)}\right]\right) - \sum_{j=i+1}^{n} \text{dual}([a_{ij}][x_i^{(l)}])\right) \Big/ \text{dual}([a_{ii}]) \right\} \\ \textbf{for } i=1 \textbf{ to } n \textbf{ do} \\ [x_i^{(l+1)}] := \left[x_i^{(l+1/2)}\right] \cap \left\{ \left([b_i] - \sum_{j=1}^{i-1} \text{dual}\left([a_{ij}]\left[x_i^{(l+1/2)}\right]\right) - \sum_{j=i+1}^{n} \text{dual}([a_{ij}][x_i^{(l+1)}])\right) \Big/ \text{dual}([a_{ii}]) \right\} \end{array}\right. \tag{9}$$

Then $\lim_{l\to\propto}[x^{(l)}] = [x^*]$.

To get an interval vector [initial]. We assume that $[A] \in \mathbb{IR}^{n\times n}$ is an H-matrix. Let $[L]$, $[U]$ *and* $[D]^{-1}$ and $[x^*]$ defined as in Theorem 2. Then we consider symmetric single step method with arbitrary $[x^{(0)}]$. We assume that $P := \langle [D]\rangle^{-1}|[L]+[U]|$ where $\rho(P) < 1$ (see Theorem 11.4 in [1]). For $m > l$ we get [2]:

$$\begin{aligned} q([x^{(m)}],[x^{(l)}]) &\le q([x^{(m)}],[x^{(m-1)}]) + \cdots + q([x^{(l+1)}],[x^{(l)}]) \\ &\le P^{m-1}\cdot q([x^{(m)}],[x^{(0)}]) + \cdots + P^{l}\cdot q([x^{(1)}],[x^{(0)}]) \\ &= P^{l}\cdot(I + P + \cdots + P^{m-l-1})\cdot P^{m-1}\cdot q([x^{(1)}],[x^{(0)}]) \\ &\le P^{l}\cdot\left(\sum_{i=0}^{\infty} P^{i}\right)\cdot P^{m-1}\cdot q([x^{(1)}],[x^{(0)}]) \\ &= P^{l}\cdot(I-P)^{-1}\cdot P^{m-1}\cdot q([x^{(1)}],[x^{(0)}]). \end{aligned}$$

Since $\lim_{l\to\propto}[x^{(l)}] = [x^*]$, it holds that (set $m := 1$)

$$q([x^*],[x^{(l)}]) \le P\cdot(I-P)^{-1}\cdot q([x^{(1)}],[x^{(0)}]) =: u,$$

Then

$$\underline{x}^{(l)} - u \le \underline{x}^*, \quad \overline{x}^* \le \overline{x}^{(l)} + u$$

Hence, we get $[x^*] \subseteq \left[\underline{x}^{(l)} - u, \overline{x}^{(l)} + u\right] := [\text{initial}]$

Algorithm 1. Interval linear systems (H-Matrix)

1. Computation of an initial interval vector

$P := \langle [D]\rangle^{-1}|[L]+[U]|$,

$[\text{initial}] := \left[\underline{x}^{(l)} - u, \overline{x}^{(l)} + u\right], \ l > 0$

2. Verification step

$[x^{(1)}] = [\text{initial}]$

repeat

if intersection = 0 **then**

Using equations (7) and (8)

else Using equation (9)

until $[x^{(l+1)}]$ *and* $[x^{(l)}]$ *are equals*

3.

if $[x^{(l+1)}]$ *and* $[x^{(l)}]$ *are equals* **then**

$\hat{x} \in [x^{(l+1)}]$ ($\hat{x}$ the exact solution)

else *no inclusion can be computed*

6. numerical examples

Example 1:

Consider

$$\begin{pmatrix} [0.4481568,0.4498432] & [0.4376422,0.4393578] \\ [0.4376938,0.4393062] & [0.6503902,0.6516098] \end{pmatrix} \cdot x = \begin{pmatrix} [0.5646710,0.5667290] \\ [0.6103170,0.6134830] \end{pmatrix}$$

Proposed method	Formal-Algebraic [3]	Hoelbig [11]
[0.9745537,1.0242698]	[0.974,1.0246]	[0.9740262,1.02468993]
[0.2460811,0.2874072]	[0.2458,0.2877]	[0.24574949,0.28783412]

Example 2:

Consider

$$\begin{pmatrix} 3 & [0,1] & [0,1] \\ [0,1] & 3 & [0,1] \\ [0,1] & [0,1] & 3 \end{pmatrix} \cdot x = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Proposed method	Hoelbig [11]
[0.111111,0.333334]	[0.04394499,0.45605500]
[0.111111,0.333334]	[0.04394499,0.45605500]
[0.111111,0.333334]	[0.04394499,0.45605500]

Example 3:

Consider

$$\begin{pmatrix} [1.6,1.8] & [0.4,0.5] & [0.2,0.3] \\ 0 & [0.6,0.8] & [0.2,0.3] \\ 0.1 & 0 & [1.4,1.8] \end{pmatrix} \cdot x = \begin{pmatrix} [1.4,2] \\ [3,3.4] \\ [2,2.3] \end{pmatrix}$$

Proposed method	Hoelbig [11]
[-1.10487815,0.33737055]	[-1.08943500,0.52626809]
[3.10433362,5.30254389]	[2.75007232,5.42039269]
[1.09236830,1.72177701]	[0.99326595,1.72943197]

Example 4: Application

we consider a linear resistive network, presented in [30].

$$\begin{pmatrix} [1.98,2.02] & [-1.01,-0.99] & 0 & 0 & 0 \\ [-1.01,-0.99] & [2.97,3.03] & [-1.01,-0.99] & 0 & 0 \\ 0 & [-1.01,-0.99] & [2.97,3.03] & [-1.01,-0.99] & 0 \\ 0 & 0 & [-1.01,-0.99] & [2.97,3.03] & [-1.01,-0.99] \\ 0 & 0 & 0 & [-1.01,-0.99] & [1.98,2.02] \end{pmatrix} \cdot v = \begin{pmatrix} 10 \\ 0 \\ 10 \\ 0 \\ 0 \end{pmatrix}$$

Proposed method	Hoelbig [11]
[6.89898004,7.29765392]	[6.88299293,7.29882524]
[3.97569664,4.405301763]	[3.95697257,4.40666378]
[5.26906122,5.656550266]	[5.25126311,5.65782779]
[2.04981395,2.3273262467]	[2.03528617,2.32835018]
[1.00461178,1.1871714692]	[0.99390173,1.18791644]

6. Conclusion

The problem of solving interval linear systems of equations is very important in practical applications.

A simple method for determining an outer solution to the linear system considered has been suggested in section 5 by using the method presented in section 4. Some numerical and practical examples are solved. The methods that presented can be applied to big real life problems such as structural engineering [18,23] without any problems.

References

[1] Alefeld, G., Herzberger,J. : Introduction to interval computations, Academic Press, London, 1983.

[2] Alefeld, G. and U. Schaefer: Iterative Methods for Linear Complementarity Problem with Interval Data. Computing 70, 2003, pp.235-259.

[3] Alexandre Goldsztejn: A Right-Preconditioning Process for the Formal-Algebraic Approach to Inner and Outer Estimation of AE-Solution Sets. Reliable Computing 11, 2005, pp.443-478 .

[4] Buckley J.J.: Solving fuzzy equations in economics and finance. Fuzzy Sets & Systems 48, 1992.

[5] Buckley J.J.: The fuzzy mathematics of finance. Fuzzy Sets & Systems 21, 1987, pp. 257-273.

[6] Chen S.H., Yang X.W.: Interval finite element method for beam structures. Finite Elements in Analysis and Design 34, 2000, pp. 75-88.

[7] J.E. Cope, B.W. Rust: Bounds on solutions of linear systems with inaccurate data SIAM J. Numer. Anal. 16 1979, pp. 950– 963.

[8] G.H. Golub, C.F. Van Loan: Matrix Computations. Johns Hopkins University Press, Baltimore, 1996.

[9] N.J. Higham: Accuracy and Stability of Numerical Algorithms. SIAM, Philadelphia, 1996.

[10] Hofschuster, W.; Kraemer, W.: C-XSC 2.0 - A C++ Class Library for Extended Scientific Computing. In: Numerical Software with Result Verification, R. Alt, A. Frommer, B. Kearfott, W. Luther (eds), Springer Lecture Notes in Computer Science 2991, 2004, pp. 15-35.

[11] Hoelbig, C.; Kraemer, W.: Self verifying Solvers for Dense Systems of Linear Equations Realized in C-XSC. Preprint 2003/1, Universitaet Wuppertal, 2003.

[12] Kaucher, E.: Ueber metrische und algebraische Eigenschaften einiger beim numerischen Rechnen auftretender Raume. PhD thesis, Karlsruhe, 1973.

[13] Kaucher, E.: Interval analysis in the Extended Interval Space $\mathbb{IR}$. Computing, Suppl. 2, 1980, pp. 33-49.

[14] Kearfott, B.: Rigorous Global Search: Continuous Problems. Kluwer Academic Publishers, The Netherlands 1996.

[15] V. Kreinovich, etl.: Optimal solution of interval linear systems is intractable (NP- hard). Interval Comput. 1, 1993, pp. 6–14.

[16] Leontief W.: Quantitative input output relations in the economic system of the United States. Review of Economics and Statistics 18, 1396, pp. 100-125.

[17] Moore R.E.: Interval Arithmetic and Automatic Error Analysis in Digital Computing. PhD thesis, Stanford University 1962.

[18] Neumaier, A.; Pownuk, A.: Linear Systems with Large Uncertainties with Applications to Truss Structures. Reliable Computing, vol. 13, issue 2, 2007, pp. 149-171.

[19] A. Neumaier: Interval Methods for Systems of Equations. Cambridge University Press, Cambridge, 1990.

[20] W. Oettli,W. Prager: Compatibility of approximate solution of linear equations with given error bounds for coefficients and right-hand sides. Numer. Math. 6, 1964, pp. 405–409.

[21] W. Oettli: On the solution set of a linear system with inaccurate coefficients . SIAM J. Numer. Anal. 2, 1965, pp. 115–118.

[22] Ortolf, H.J.: Eine Verallgemeinerung der Itervallarithmetik. Geselschaft fuer Mathematik und Datenverarbeitung, Bonn 11, 1969, pp. 1-71.

[23] Popova, E.; Datcheva, M.; Iankov, R.; Schanz, T.: Mechanical Models with Interval Parameters . In: K. Guerlebeck, L. Hempel, C. Koenke (Eds.) IKM2003: Digital Proceedings of 16th International Conference on the Applications of Computer Science and Mathematics in Architecture and Civil Engineering, ISSN 1611-4086, Bauhaus University Weimar, 2003.

[24] Popova, E.: Multiplication distributivity of proper and improper intervals. Reliable Computing, vol. 7, issue 2, 2001, pp. 129-140.

[25] J. Rohn: Overestimations in bounding solutions of perturbed linear equations. Linear Algebra Appl. 262, 1997, pp. 55–65.

[26] B.W. Rust,W.R. Burrus: Mathematical Programming and the Numerical Solution of Linear Equations. Elsevier, New York, 1972.

[27] S.P. Shary: On optimal solution of interval linear equations. SIAM J. Numer. Anal. 32, 1995, pp. 610–630.

[28] Varga, R. S.: Matrix iterative Analysis. Prentice-Hall, Englewood Cliffs, New Jersey, 1962.

[29] Wu C.C., Chang N.B.: Grey input output analysis and its application for environmental cost allocation. European Journal of Operational Research 145, 2003, pp. 175-201.

[30] OKumura, K.: An Application of Interval Operations to Electric Network Analysis. Bull. of the Japan Soc. for Industrial \ Applied Mathematics, 2, 1993, pp. 115-127.

Hassan Badry Mohamed A. El-Owny is Assistant Professor of Computer Science at Faculty of Science, Aswan University, Aswan, Egypt. His current position is Assistant Proffersor at Taif University, Taif, KSA. He received the PhD degree in ``Computer Science" at University of Wuppertal(Germany). He is an Aboard Editor of several international journals. He has published extensively in internationally refereed journals and conferences. His main research interests are: Scientific Computing, Computer Algebra, Interval Arithmatics, Parametric systems, Approximation of Uncertain Data and Software Tools with Result Verification.

Artificial Neural Networks in Medical Images for Diagnosis Heart Valve Diseases

Atta Elalfi[1], Mohamed Eisa[2] and Hosnia Ahmed[3]

**[1] Department of Computer Science, Mansoura University
Egypt**

**[2] Department of Computer Science, Port-Said University
42526 Port Said, Egypt**

**[3]Department of Computer Science, Mansoura University
Egypt**

Abstract

Neural networks are currently a hot research area in medicine. Medical image recognition algorithms have been widely applied to help with the diagnosis of various diseases more accurately. This paper presents an image processing-based artificial neural network for the diagnosis of heart valve diseases. The goal of the paper is to implement image processing techniques by extracting texture features from medical echocardiography images, combining intensity histogram features and Gray Level Co-occurrence Matrix (GLCM) features, then developing an artificial neural network for automatic classification based on back-propagation algorithm to classify heart valve diseases more accurately. The proposed method performance was evaluated in terms of precision, recall and accuracy. The experimental results confirm the efficiency of the proposed method that provides good classification efficiency.

***Keywords**: Computer-Aided Diagnosis (CAD), Neural networks, Texture Features, Intensity Histogram Features, GLCM Features, Back-Propagation Classifier, Heart Valve Diseases.*

1. Introduction

Medical and healthcare sector are a big industry directly related to every citizen's quality of life. Image based medical diagnosis is one of the important service areas in this sector [1].

In recent years, considerable efforts have been made in CAD using medical images to improve a clinician's confidence in the analysis of medical images. Evaluation of medical images by a clinician is qualitative in nature and may vary from person to person. A lot of research efforts have been directed to the field of medical image analysis with the aim to assist in diagnosis and clinical studies [2].

The number of medical images has grown significantly in the recent years. These images are very important for clinical diagnosis, localization of pathology, study of anatomical structure, treatment planning, evolution of therapy, computer integrated surgery, surgical planning, post-surgical assessment and abnormality detection [3].

Researches showed that the most human deaths in the world are due to heart diseases. Heart valve disorders are of importance among the heart diseases. For this reason, early detection of heart valve disorders is one of the most important medical research areas [4].

Nowadays, cardiologists have access to diverse techniques such as electrocardiograms, chest X-rays, ultrasound imaging, doppler techniques, angiography and transesophageal echocardiograph to better inspect and scrutinize the functionality of heart [5].

Echocardiography is a common clinical procedure for diagnosing heart diseases, especially valve ones. When digital echocardiographies are available, computer-aided diagnosis may help physicians in having a more accurate decision [6].

Various artificial intelligence techniques such as artificial neural network and fuzzy logic are used for classification problems in the area of medical diagnosis [7].

Most of these computer-based systems are designed by using artificial neural network techniques [8].

Artificial Neural Networks (ANNs) are one of the popular methods for classification problems compared to most traditional classification approaches. ANNs are nonlinear, nonparametric, and adaptive. They can theoretically approximate any fundamental relationship with arbitrary accuracy [9].

There are four reasons to use ANN as a classifier: (i) weights representing the solution are found by iterative training, (ii) it has a simple structure for physical implementation, (iii) it can easily map complex class distributions, and (iv) generalization property of the ANN produces appropriate results for the input vectors that are not present in the training set [10].

ANNs have been of increasing interest in medical image processing [1, 2].

In summary, the applications of ANNs in medical image processing have to be analyzed individually, although many successful models have been reported in the literature. ANN has been applied to medical images to deal with the issues that cannot be addressed by traditional image processing algorithms or by other classification techniques. By introducing artificial neural networks, algorithms developed for medical image processing and analysis often become more intelligent than conventional techniques [11].

The structure of this paper is as follows: section 2 presents the framework of the proposed method, the algorithm used in this system is commonly divided into three stages: preprocessing, feature extraction and Back-Propagation Classifier. Section 3 describes the process of designing the ANN. Section 4 is the experimental results, followed by conclusions at section 5.

2. Proposed method

The proposed method comprises three stages: preprocessing, feature extraction and back-propagation as classifier. Figure 1 shows the block diagram of the proposed Back-Propagation Artificial Neural Network (B-PANN) method.

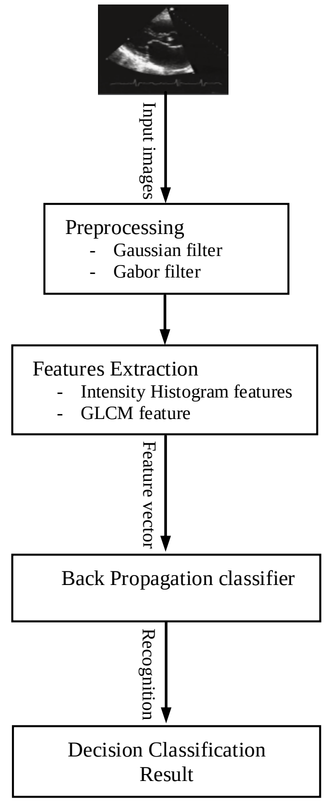


Fig.1. Block diagram of the proposed B-PANN

The proposed method is detailed in Figure 2

Input: a set of medical echocardiography images

Output: classification of the echocardiography images for each disease

Begin

Step1: preprocessing phase (image processing)

- convert images to gray scale
- resize the images
- reducing noise by applying Gaussian filter
- images enhancement with the application of Gabor filter

Step 2: feature extractions

- compute the intensity histogram features: Mean, Standard deviation, Skewness, Kurtosis, Energy, Entropy.
- compute GLCM features: Contrast, Correlation, Entropy, Homogeneity.

Step 3: Determine the structure of B-PANN

- Number of input neurons
- Number of output neurons

- learning rate
- Number of hidden neurons
- Number of epochs
- Activation function

Step 4: classification phase(training and testing with Back-Propagation classifier)

Input:

A set of input features of echocardiography images to be classified

Processing (Back-Propagation Algorithm)

Output:

The final classification output

End

Fig .2 the proposed algorithm

2.1. Preprocessing

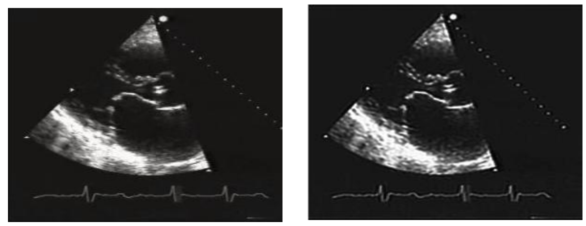
Image preprocessing refers to the initial processing of raw image to correct geometric distortions, calibrate data radio metrically and eliminate the noise and clouds in the data [12]. A preprocessing phase of the images is necessary to improve the quality of the images and make the feature extraction phase more reliable. Preprocessing is always a necessity whenever the data to be mined are noisy, inconsistent or incomplete. Preprocessing significantly improves the effectiveness of the pattern recognition techniques [13].

One of the most important stages in medical images detection and analysis is the use of image enhancement techniques which improve the quality (clarity) of images for human viewing. Removing blurring and noise, increasing contrast and revealing details are examples of enhancement operations [14].

This section introduces the preprocessing techniques on the images before the features extraction phase. The following two methods (Gaussian and Gabor filters) are used for making the image better and enhance it from noising corruption to interference.

2.1.1 Gaussian Filter

The medical images contain a large area of background noise, which is useless in medical diagnosis. Gaussian filter is group of low-pass filters which passes over low-frequency components and reduces high-frequency components [12]. Gaussian filter is a filter whose impulse response is Gaussian function. Gaussian filters are designed to give no overshoot to a step function input while minimizing the rise and fall time. This behavior is closely connected to the fact that the Gaussian filter has the minimum possible group delay [15]. The proposed method reduces the noise by applying the Gaussian filter in the preprocessing stage. Figure 3 describes (a) the original image and (b) reducing noise by Gaussian filter.



y

(a) Original image (b) Gaussian filter

Fig.3. the result of applying Gaussian filter

2.1.2 Gabor Filter

Gabor function has been recognized as a very useful tool in computer vision and image processing, especially for texture analysis, due to its optimal localization properties in both spatial and frequency domains [14].

Gabor filter has been successfully applied to the fields of image processing and image analysis, including edge detection, texture image segmentation/discrimination, texture classification/recognition and image enhancement [16, 17].

Gabor filter is a linear filter whose impulse response is defined by a harmonic function multiplied by the Gaussian function [18]. Because of the multiplication-convolution property, the Fourier transform of the Gabor filter's impulse response is the convolution of the Fourier transform of the harmonic function and the Fourier transform of the Gaussian function [12, 19].

Gabor filter has been extended to 2D operation. A 2D Gabor filter is an oriented complex sinusoidal grating modulated by a 2D Gaussian function [16, 17, 20]

$$h(x,y) = g(x,y)\exp[2\pi j(Ux+Vy)] = h_R(x,y) + jh_I(x,y) \quad (1)$$

Where (U,V) is a single spatial frequency, $g(x,y)$ is the Gaussian function with scale parameter σ, and $h_R(x,y)$ and $h_I(x,y)$ are the real and imaginary parts of $h(x,y)$, respectively

$$g(x,y) = \frac{1}{2\pi\sigma^2}\exp\left\{-\frac{x^2+y^2}{2\sigma^2}\right\} \quad (2)$$

Gabor filter is a band pass filter centered on frequency (U,V), with a bandwidth determined by σ. The parameters of the Gabor filter are represented by the spatial frequency U, V and the scale σ. Usually, a radial frequency

$f=\sqrt{U^2+V^2}$, with orientation θ = $\tan^{-1}$(V/U), are used in polar coordinates to specify the filter (f, θ, σ). Gabor filtered output of an image *i(x,y)* is obtained by the convolution of the image with the specified Gabor function. The local energy measure at a point *(x,y)* is defined as [16].

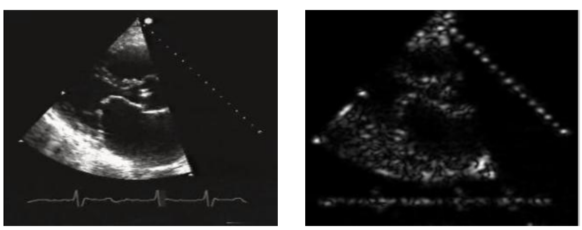
$$E(x,y|f,\theta,\sigma)=C_R^2(x,y|f,\theta,\sigma)+C_1^2(x,y|f,\theta,\sigma) \quad (3)$$

where

$$C_R(x,y|f,\theta,\sigma)=\sum_{I=-w}^{w}\sum_{m=-w}^{w} i(X+l,y+m)h_R(l,m)$$

$$\text{and } C_1(x,y|f,\theta,\sigma)=\sum_{I=-w}^{w}\sum_{m=-w}^{w} i(X+l,y+m)h_1(l,m) \quad (4)$$

represent the discrete convolution of the real and imaginary components of *h(x,y)* with the image over a given neighborhood with a fixed window size of M = 2w +1. The resulting feature image, *E(x,y),* contains a distribution of local energy measures, which depend strongly on the choice of the design parameters (f, θ, σ) of the single Gabor filter [16]. Figure 4 describes (a) the original image and (b) the enhanced image using Gabor Filter.



(a) Original image (b) Enhanced by Gabor

Fig.4 the results of applying Gabor enhancement technique

2.2. Features Extraction

During the process of developing the CAD systems, feature extraction is one of the most important steps for recognizing abnormal regions from the medical image [21]. Image feature extraction is an important step in medical images classification. These features are extracted using image processing techniques [7]. The texture feature was used in various areas including face detection, face recognition, medical research and satellite image analysis [22]. Texture features have been utilized for many medical image applications [23].

Texture feature is one of most important feature analysis methods in CAD systems for disease diagnosis [19]. Texture features such as intensity histogram, GLCM, Gray Level Difference Method (GLDM), Gray Level Run-Length Method (GLRLM) and the Spatial Gray Level Dependent Method (SGLDM) have been widely used to represent medical image characteristics that are inaccessible to human observers [21, 24]. Intensity histogram features and GLCM features are extracted in our paper.

2.2.1. Intensity Histogram Features

In general, any image processing and analysis application would require particular features for classification /segmentation. Mainly texture features and statistical features are of more significant in pattern recognition. A frequently used approach for texture analysis is based on statistical properties of intensity histogram. One such measure is based on statistical moments [25, 10]. One of the simplest ways to extract statistical features in an image is to use the first-order probability distribution of the amplitude of the quantized image. They are generally easy to compute and largely heuristic [26, 23]. The first order histogram estimate of P(b) is simply

$$P(b)=\frac{N(b)}{M} \quad (5)$$

where b is a gray level in an image, M represents the total number of pixels in a neighborhood window of specified size centered around the pixel, and N(b) is the number of pixels of gray value b in the window where 0 ≤ b ≤ L -1 [26]. The following table shows texture features extracted based on the intensity histogram features in this work.

Table 1. Some texture features extracted based on the intensity histogram features [26].

Features	***Equation***
Mean	$S_M=\bar{b}=\sum_{b=0}^{L-1} bP$ (6)
Standard deviation	$S_D=\sigma_b=\left[\sum_{b=0}^{L-1}(b-\bar{b})^2 p(b)\right]^{1/2}$ (7)
Skewness	$S_S=\frac{1}{\sigma_b^3}\sum_{b=0}^{L-1}(b-\bar{b})^3 P(b)$ (8)
Kurtosis	$S_K=\frac{1}{\sigma_b^4}\sum_{b=0}^{L-1}(b-\bar{b})^4 P(b)-3$ (9)
Energy	$S_N=\sum_{b=0}^{L-1}[P(b)]^2$ (10)
Entropy	$S_E=\sum_{b=0}^{L-1}P(b)\log_2\{P(b)\}$ (11)

2.2.2. GLCM Features

GLCM is a second-order method for measuring gray level textures in an image [27].

GLCM is frequently used method for medical image analysis, facial analysis and classification. This method provides information regarding the relative position of two pixels with respect to each other [22, 28]. The following table shows texture feature extracted from GLCM in this work.

Table 2. some texture features extracted from gray level co-occurrence matrices [29].

Features	Equation
Homogeneity	$H = \sum_{i=0}^{G-1} \sum_{i=0}^{G-1} \{P(i,j)\}^2$ (12)
Contrast	$C = \sum_{n=0}^{G-1} \left\{ n^2 \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i,j) \right\}, \lvert i-j \rvert = n$ (13)
Entropy	$E = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i,j) \log\{P(i,j)\}$ (14)
Correlation	$O = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{ijP(i,j) - (m_i m_j)}{\sigma_i \sigma_j}$ (15)

Where i and j: row and column numbers in the GLCM matrix, m_i and σ_i are the mean and standard deviation of P(i,j) rows, and m_j and σ_j the mean and standard deviation of P(i, j) columns, respectively.

2.3 The Back-Propagation Classifier

One of the most common ANN used for classifications is the feed-forward network. In a feed-forward network, the neurons in each layer are only connected with the neurons in the next layer. Feed-forward networks commonly use the back-propagation (B-P) supervised learning algorithm to dynamically alter the weight and bias values for each neuron in the network [11]. A back-propagation neural network was created to classify the images is presented in [30].

2.3.1. The architecture of B-P ANN

In this paper, a neural network to perform image classification is constructed as follows; it has three layers; input, hidden and output layer. The number of neurons in the input layer is determined by the number of features selected. The number of neurons in the output layer is determined by the number of classes represented in the network. The number of hidden neurons is determined based on experiments, starting with a single node and increasing the number of nodes until the highest performance was found [31, 32, 33].

The architecture of the neural network is illustrated in the following Figure.

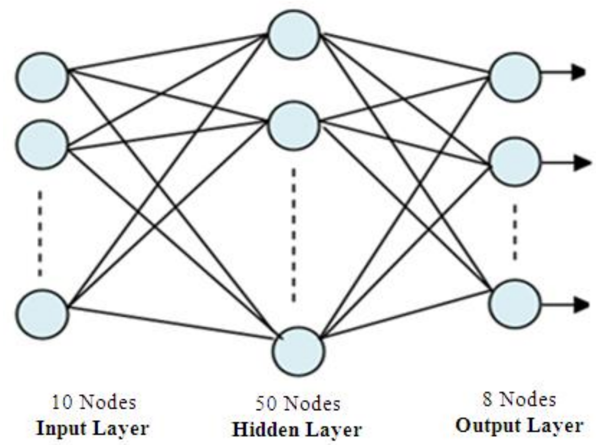


Fig.5. the architecture of B-P ANN

The following table shows the architecture and parameters of the designed B-P ANN.

Table.3 The architecture and parameters of the designed B-P ANN

ANN architecture:	
Number of input neurons	10
Number of hidden neurons in hidden layer	50
Number of output neurons	8
Activation function	Log-sigmoid
ANN training parameters	
Learning rule	Back-propagation
Learning rate	0.1
Number of epochs	3000
Error rate measure	MSE(Mean squared error)

3. Training and Testing

A training stage consists of five major steps:

1. Load an image from a database.
2. Pre-process and extract the features of the image by the selected features discussed earlier.
3. Applying features (feature vector) an input to B-P Algorithm.
4. Training the back- propagation neural network.
5. Save as .mat file.

A Testing stage consists of five major steps:

1. Loading of an image to be tested from a database.

2. Pre-process and extracting the features of the image by the feature extraction techniques discussed earlier.
3. Application of extracted features as input to a trained neural network.
4. B-P algorithm makes decision according to value of features.
5. Checking output generated from a neural network.

Figure 6 shows the training and testing stages of B-PANN.

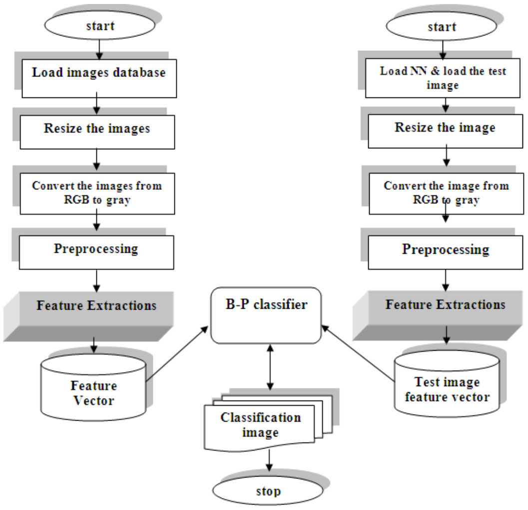


Fig. 6 Training & Testing stages of B-P ANN

4. Experimental Results

The proposed method has been implemented using MATLAB. The image database in the experiment is provided by Mansoura University Hospitals, several radiological center and CDs by the web. Our database contains 120 medical echocardiography images. These images are organized in 8 classes of 15 images: Aortic Regurge (AR), Aortic Stenosis (AS), Mitral Regurge (MR), Mitral Stenosis (MS), Pulmonary Regurge (PR), Pulmonary Stenosis (PS), Tricusped Regurge (TR) and Tricuspid Stenosis (TS). In order to evaluate the performance of this work, precision, recall and Accuracy were used. These measures are defined as: [34, 35]

$$\text{Accuracy} = \frac{\text{Number of Corrrctly Classified Testing Samples}}{\text{Total Number of Testing Samples}} \times 100 \qquad (16)$$

$$Precision(P) = \frac{\text{True positives}}{\text{True positives} + \text{False positives}} \quad 0 \le P \le 1 \qquad (17)$$

$$\text{Re}\,call\,(R) \frac{\text{True positives}}{\text{True positives} + \text{False negatives}} \quad 0 \le R \le 1 \qquad (18)$$

where True positives: correctly identified.

False positives: incorrectly identified
False negative: incorrectly denied

Recall measures the proportion of the positive examples that are correctly identified while precision measures the proportion of the nominated positive examples that are correct [35]. Precision is the fraction of the number of true positive predictions divided by the total number of true positives and false positives in the set. Recall is the fraction of the number of true positive predictions divided by the total number of true positives and false negatives in the set.

In the experiment, 80 echocardiography images of heart valve diseases were used as test images. The proposed method was used to distinguish eight different heart valve diseases through automatic recognition. The results were then compared with the results of manual interpretation by medical professionals to determine the accuracy rate of the proposed scheme.

Performance of B-P classifier

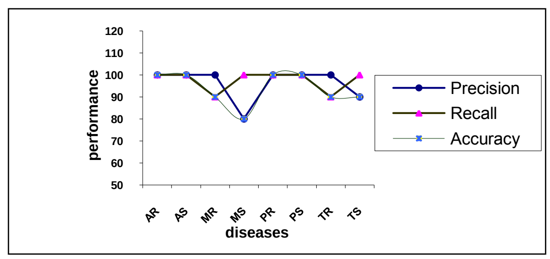


Fig.7. the relation between precision, recall and accuracy.

The experimental results shown in Figure 7 prove that the proposed scheme is capable of automatically recognizing different heart valve diseases in echocardiography images by using the B-P classifier, as the accuracy rate was a perfect 1.00 for AR, AS, PR, PS. As for MR, TR, TS the accuracy rate was 0.90, and the MS, the accuracy rate was 0.80. Altogether, the total accuracy rate was 93.75%, The experimental results confirm the efficiency of the proposed method in recognizing valve heart diseases.

In another method to evaluate the performance of this work, precision and recall were used [36].

Precision: is the fraction of the relevant images which has been retrieved (from all retrieved):

Precision=A/B

Where, A is "Relevant retrieved" and
B is "All Retrieved images"

Recall: is the fraction of the relevant images which has been retrieved (from all relevant):

Recall=A/D

Where, A is "Relevant retrieved" and
B is "All Retrieved images in Database"

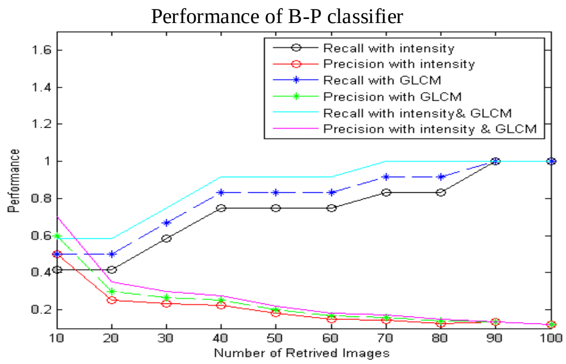


Fig.8 the relation between precision /recall v_s number of retrieved images

A retrieved image is considered as a correct match if it is the same category as the query image. The results for the query image are compared with database images. We then count how many of these belong to the correct category and define the retrieval rate. The higher value of both precision and recall shows better performance of the method.

In this work, we compare the performance of individual classifier with different features: intensity histogram features, GLCM features, and combined features (intensity histogram and GLCM).

The average retrieval precision / recall vs. number of retrieved images curves for intensity histogram features, GLCM features and combined intensity histogram and GLCM features are plotted in Figure 8. It can be seen from Figure 8 that the combined method (features) achieves higher performance in terms of the retrieval precision and recall than with applying intensity histogram features and GLCM features.

5. Conclusions

In this paper, approach based on image processing and neural network technology using feed forward neural network trained by the error back-propagation algorithm that allowed its use to classify heart valve diseases is proposed. The preprocessing techniques applied to the images are the Gaussian filter and the Gabor filter, the proposed method of feature extraction is the combination of intensity histogram features and GLCM features. All these features were fed as input to the B-P ANN which used to identify the various heart valve diseases.

The performances of the classification algorithm are evaluated using precision-recall and accuracy rate. The results demonstrate the effectiveness of the proposed algorithm.

With the help of the proposed method, beginner physicians and other professionals who work in the field of heart diseases will be able to classify heart valve diseases and give diagnosis of heart valve diseases with high accuracy and efficiency. For hospitals and other medical centers, this means a great relief on the demand of human resources.

References

[1] R. Suganya, and S. Rajaram, "Content Based Image Retrieval of Ultrasound Liver Diseases Based on Hybrid Approach", American Journal of Applied Sciences 9 (6), 2009, pp. 938-945.
[2] N. Sharma, A. K. Ray, S. Sharma, K. K. Shukla, S. Pradhan and L. M. Aggarwal, "segmentation and classification of medical images using texture-primitive features: Application of BAM-type artificial neural network", journal of medical physics, 33(3), 2008, pp.119-126.
[3] A. Rajaei, L. Rangarajan, "Medical Images Segmentation Using Linear Combination of Gabor Filtered Images", International Journal of Machine Intelligence (IJMI), Vol. 3, Issue. 4, 2011, pp. 212-216.
[4] I. T. A. Arslan, E. IlkayK, "An expert system for diagnosis of the heart valve diseases", Expert Systems with Applications (23), 2002, pp. 229–236.
[5] S. Babaei, A. Geranmayeh, "Heart sound reproduction based on neural network classification of cardiac valve disorders using wavelet transforms of PCG signals", Computers in Biology and Medicine (39), 2009, pp. 8–15.
[6] K. Chykeyuk, D. A. Clifton, J. Alison Noble, "Feature Extraction and Wall Motion Classification of 2D Stress Echocardiography with Relevance Vector Machines", IEEE, 2011, pp. 677-680, 2011.
[7] R. Nithya, B. Santhi, "Comparative Study on Feature Extraction Method for Breast Cancer Classification Journal of Theoretical and Applied Information Technology", Vol. 33 No. 2, 2011, pp. 220-226.
[8] N. Acır, "A support vector machine classifier algorithm based on a perturbation method and its application to ECG beat recognition systems", Expert Systems with Applications (31), 2006, pp. 150–158.
[9] J. Lan , M. Y. Hu , E. Patuwo, G. Peter Zhang, "An investigation of neural network classifiers with unequal misclassification costs and group sizes", Decision Support Systems (48), 2010, pp. 582–591.
[10] Z. Iscan, A. Yuksel, Z. Dokur, M. Korurek, T. Olmez,"Medical image segmentation with transform and moment based features and incremental supervised neural network", Digital Signal Processing (19), 2009, pp. 890–901.
[11] J. Jiang, P. Trundle, J. Ren, "Medical image analysis with artificial neural networks", Computerized Medical Imaging and Graphics (34), 2010, pp. 617–63.
[12] M. N. S. Zainudin, M. M. Said and M. M. Ismail, "Feature Extraction on Medical Image using 2D Gabor filter, Applied Mechanics and Materials", Vol. 52-54, 2011, pp. 2128-2132.
[13] M. L. Antoni, O. R.Zaiane, A. Coman, "Application of Data Mining Techniqqes for Medical Images Classification", Proceedings of the second international workshop on multimedia data mining (MDM/KDD), in conjunction with ACM SIGKDD conference. San Francisco, USA, August 26, 2001, pp. 94-101.
[14] S. Vishukumar, K. Patel and P. Shrivastava, " Implementation of Medical Image Enhancement Technique using Gabor Filter", International Journal of Current Engineering Technology, Vol. 2, No. 2, 2012, PP. 251-255.

[15] P. S. Hiremath, "Removable of Gaussian Noise in Despeckling Medical Ultrasound Images", the International Journal of Computer science & Applications (TIJCSA), Vol. 1, No. 5, 2012, pp. 25-35.

[16] Ma Li, R. C. Staunton, "Optimum Gabor filter design and local binary patterns for texture segmentation", Pattern Recognition Letters 29, 2008, pp. 664–672.

[17] A. Bodnarova, M. Bennamoun, S. Latham, "Optimal Gabor filters for textile flaw detection", Pattern Recognition (35), 2002, pp. 2973 – 2991.

[18] J. F. Khan, R. R. Adhami, S. M.A. Bhuiyan, "A customized Gabor filter for unsupervised color image segmentation", Image and Vision Computing (27), 2009, pp. 489–501.

[19] M. S. AL-TARAWNEH, "Lung Cancer Detection Using Image Processing Techniques", Leonardo Electronic Journal of Practices and Technologies, Issue. 20, 2012, pp. 147-158.

[20] F. Bianconi and A. Fernandez, "Evaluation of the effects of Gabor filter parameters on texture classification", Pattern Recognition (40), 2007, pp. 3325–3335.

[21] S. Hu Peng, D. HwanKim, N. Seok-LyongLee, M. KwanLim, "Texture feature extraction based on a uniformity estimation method for local brightness and structure in chest CT images", Computers in Biology and Medicine (40), 2010, pp. 931-942.

[22] U. Akilandeswari, R. Nithya, B. Santhi, "Review on feature extraction methods in pattern classification", European Journal of Scientific Research, Vol. 71, No. 2,2012, pp. 265-272.

[23] I. Christoyianni, A. Koutras, E. Dermatas, G. Kokkinakis, "Computer aided diagnosis of breast cancer in digitized mammograms", Computerized Medical Imaging and Graphics (26) ,2002, pp. 309–319.

[24] M. Radhakrishnan1 and T. Kuttiannan, "Comparative Analysis of Feature Extraction Methods for the Classification of Prostate Cancer from TRUS Medical Images", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue. 1, No. 2, 2012, PP. 171-179.

[25] H. S. Sheshadri and A. Kandaswamy, "Breast Tissue Classification Using Statistical Feature Extraction of Mammograms", Vol. 23, No. 3, 2006, pp. 105-107.

[26] E. H. Aria, M. R. Saradjian, J. Amini, C. Lucas," Generalized Co-occurrence Matrix to Classify IRS-1D Images using Neural Network", Faculty of Engineering, University of Tehran, Tehran, Iran, Available at: http://www.isprs.org/proceedings.

[27] P. W. Mirowski, D. M. Tetzlaff, "Retrieving scale from quasi-stationary images", Pattern Recognition Letters (29), 2008, pp. 754–767.

[28] M. Vasantha, V. S. bharathi, R. Dhamodharan, "Medical Image Feature, Extraction, Selection And Classification", International journal of Engineering Science and Technology, Vol. 2, No. 6, 2010, pp. 2071-2076.

[29] J. F. Vargas, M. A. Ferrer, C. M. Travieso, J. B. Alonso, " offline signature verification based on gray level information using texture features", Pattern Recognition (44), 2011, pp. 375-385.

[30] M. Veluchamy, K. Perumal and T. Ponuchamy, "Feature Extraction and Classification of Blood Cells Using Artificial Neural Network", American Journal of Applied Sciences 9(5), 2012, PP. 615-619.

[31] K. Roy, C. Chaudhuri, M. Kundu, M. Nasipuri, D. K. Basu, "Comparison of the Multi Layer Perceptron and the Nearest Neighbor Classifier for Handwritten Numeral Recognition", Journal of Information Science and Engineering (21), 2005, pp. 1247-1259.

[32] G. Castellano and A. M. Fanelli, "Variable selection using neural-network models", Neurocomputing (31), 2000, pp. 1-13.

[33] D.Ansari,J.Nilsson,R.Andersson,S.Regner,B.Tingstedt, B. Andersson, "Artificial neural networks predict survival from pancreatic cancer after radical surgery", The American Journal of Surgery, (205), 2013, pp. 1-7.

[34] A. Rajaei, L. Rangarajan, "Wavelet Features Extraction for Medical Image Classification" An International Journal of Engineering Sciences, Vol. 4, 2011, pp. 131-141.

[35] I. Middleton, R. I. Damper, "Segmentation of magnetic resonance images using a combination of neural networks and active contour models", Medical Engineering & Physics (26), 2004, pp. 71–86.

[36] H. B. Kekre and K. Sonawane, "Statistical Parameters Based Feature Extraction Using Bins With Polynomial Transform of Histogram", International Journal of Advances in Engineering & Technology, Vol. 4, Issue. 1, 2012, PP. 510-524.

Proposed Software Testing Using Intelligent techniques (Intelligent Water Drop (IWD) and Ant Colony Optimization Algorithm (ACO))

Laheeb M. Alzubaidy [1], Baraa S. Alhafid [2]

[1] **Software Engineering, Mosul University, Collage of Computer sc. And Mathematic**
Mosul , Iraq

[2] **Software Engineering, Mosul University, Collage of Computer sc. And Mathematic**
Mosul , Iraq

Abstract

This paper proposed software testing system by using artificial intelligent techniques. And that was conducted through Suggestion Intelligent Water Drop Algorithm (IWD) with white box testing for generated basis bath testing and using Ant Colony Optimization Algorithm (ACO) for test data generation. Correctly generated Test data helps in reducing the effort while testing the software. Automatic generation of test data is required to enables the corporation which develops the program to save time and costs as well as ensuring the test process quality, which is estimated by 50% of the product cost.

Keywords*: Artificial Intelligent Techniques, Intelligent Water Drop (IWD), Ant Colony Optimization (ACO), Software Testing, Path Basis Testing, Test Data Generation***.**

1. Introduction

Software test is the main approach to find errors and defects assuring the quality of software. Software testing is an expensive component of software development and maintenance. Testing is a complex, labor-intensive, and time consuming task that accounts for approximately 50% of the cost of a software system development [16]. Aim of the software testing is to uncover errors and faults present in the program, so that customer requirement can be properly fulfilled. Testing phase includes in the review of specification, analysis, design, and implementation part of the Software Development Life Cycle (SDLC). Manual generation of test data for testing the program, results in low reliability and high cost [12]. Due to the lack of cost and reliability, automation of testing process is necessary, so that the cost of testing can be reduced. Artificial Intelligence (AI) based techniques can help in removing this situation. AI based technique helps in solving the problem by using fast and proper judgments rather than using step by step deduction [2].

The paper is structured as follows: section 2 introduces related work, section 3 present testing in software engineering with the objective and type of software testing, section 4 describes the intelligent water drop algorithm, section 5 describes the ant colony optimization algorithm, section 6 includes the proposed work, section 7 includes Conclusions.

2. Related work

Various techniques have been proposed for automated testing to reduce efforts to a remarkable extent.

Andreas W., Stefan W., Joachim W.in (2007) suggested an empirical comparison of a genetic algorithm and a particle swarm algorithm applied to evolutionary structural testing. They selected 25 artificial test objects that cover a broad variety of search space characteristics (e.g. varying number of local optima), and 13 industrial test objects taken from various development project. The results indicate that particle swarm optimization is well-suited as a search engine for evolutionary structural testing and tends to outperform genetic algorithms in terms of code coverage achieved by the delivered test cases and the number of needed evaluations [1].

Praveen R. S., Tai-hoon K.in (2009) presents a method for optimizing software testing efficiency by identifying the most critical path clusters in a program. They do this by developing variable length Genetic Algorithms that optimize and select the software path clusters which are weighted in accordance with the criticality of the path. Exhaustive software testing is rarely possible because it becomes intractable for even medium sized software. Typically only parts of a program can be tested, but these parts are not necessarily the most error prone. Therefore, they are developing a more selective approach to testing by focusing on those parts that are most critical so that these paths can be tested first. By identifying the most critical paths, the testing efficiency can be increased [11].

Surender S. D., Jitender K. C. , Shakti K. in (2010) presents an artificial bee colony based novel search technique for automatic generation of structural software tests. Test cases are symbolically generated by measuring fitness of individuals with the help of branch distance based objective function. Evaluation of the test generator was performed using ten real world programs. Some of these programs had large ranges for input variables. Results show that the new technique is a reasonable alternative for test data generation, but doesn't perform very well for large inputs and where constraints are having many equality constraints [15].

Sanjay S., Dharminder K., H M Rai and Priti S. in (2011) presents a technique that based on a combination of genetic algorithm (GA) and particle swarm optimization (PSO), and is thus called GPSCA (Genetic-Particle Swarm Combined Algorithm) which is used to generate automatic test data for data flow coverage with using dominance concept between two nodes. The performance of the proposed approach is analyzed on a number of programs having different size and complexity. Finally, the performance of GPSCA is compared to both GA and PSO for generation of automatic test cases to demonstrate its superiority [16].

3. Testing in software engineering

There are a many definitions of software testing, but one can shortly define that as: "**A process of executing a program with the goal of finding errors**". So, testing means that one inspects behavior of a program on a finite set of test cases (a set of inputs, execution preconditions, and expected outcomes developed for a particular objective [9].

The objectives of software testing are: [13]

- A good test case is one that has a high probability of finding an as-yet undiscovered error.
- A successful test is one that uncovers an as-yet undiscovered error.
- Testing is a process of executing a program with the intent of finding an error.

The Testing type in software engineering is:

- Black box Testing
- White box Testing
- Gray box Testing

3.1 White box Testing

In this paper a White box testing is used, White box testing based on an analysis of internal working and structure of a piece of software. White box testing is the process of giving the input to the system and checking how the system processes that input to generate the required output as illustrated in Fig 1 .It is necessary for a tester to have the full knowledge of the source code. White box testing is applicable at integration, unit and system levels of the software testing process. In white box testing one can be sure that all parts through the test objects are properly executed [10].

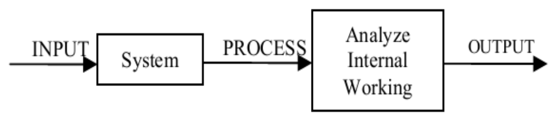


Fig. 1 Represent working process of White Box Testing

The types of white box testing techniques are : [10]

- Control Flow Testing
- Branch Testing
- Basis Path Testing
- Data Flow Testing
- Loop Testing

3.1.1 Basis Path Testing

Basis path testing is a white-box testing technique first proposed by Tom McCabe [13] and it allows the test case designer to produce a logical complexity measure of procedural design and use this measure as an approach for outlining a basic set of execution path (basic set is the set of all the execution of a procedure) These are test cases that exercise basic set will execute every statement at least once. Basic path testing makes sure that each independent path through the code is taken in a predetermined order. For this reason Basis Path Testing is used in this paper.
The method devised by McCabe to carry out basis path testing has four Steps. These are [5]:

- Compute the program graph.
- Calculate the cyclomatic complexity.
- Select a basis set of paths.
- Generate test cases for each of these paths

a. Flow Graph Notation

Before we consider the basis path method, a simple notation for the representation of control flow called allow graph (or program graph) must be introduced, The flow graph depicts logical control flow using the notation illustrated in Fig 2, Each structured construct has a corresponding flow graph symbol [13].

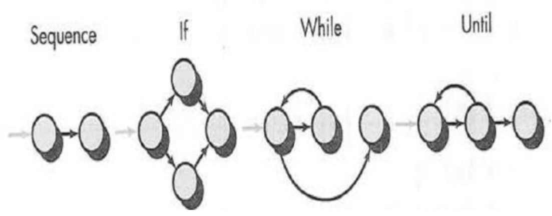


Fig 2 Flow Graph Notation.

Control Flow Graph (CFG) describes the sequence in which the statements/instructions of a program are executed. It is representation of flow of control through the program. CFG is directed graph in which each node is a program statement/basic block and each edge represents the flow of control between statement/basic blocks. A basic block is a sequence of consecutive statements in which flow of control enters at the beginning and leaves at the end without halt or possibly of branching except at the end [8]. In a CFG, a node including condition is called a predicate node as shown in Fig 3, and edges from the predicate node must converge at a certain node. Area defined by edges and nodes is referred to as region [13].

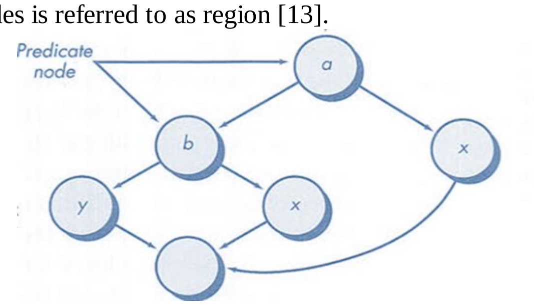


Fig 3 Predicate node.

On a flow Graph as shown in Fig 4 :

- the symbol arrows called as Edges that represent the flow of control
- Circles are called as nodes, which represent one or more actions.
- Areas bounded by edges and nodes called regions

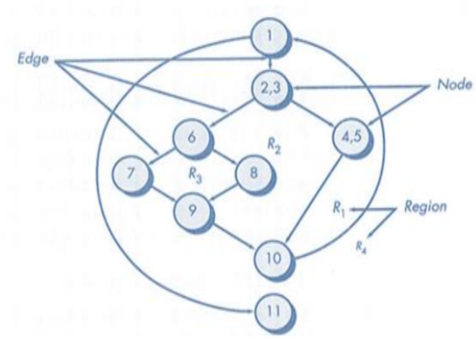


Fig4 Flow Graph .

b. Cyclomatic Complexity (CC)

The notion of Cyclomatic complexity was presented by McCabe. Cyclomatic complexity is software metric that delivers a quantitative degree of the logical difficulty of a program. Cyclomatic Complexity (CYC) is derived as the number of edges of the program's control-flow graph minus the number of its nodes plus two times the number of its linked components. Cyclomatic complexity purely depends on the Control Flow Graph (CFG) of the program to be tested [14] complexity is computed in one of three ways [13]:

- The number of regions of the flow graph corresponds to the Cyclomatlc complexity.
- Cyclomatic complexity V(G) for a flow graph G is defined as V(G)=E-N+2

Where E is the number of flow graph edges and N is the number of flow graph nodes.

- Cyclomatlc complexity V(G) for a flow graph G is also defined as v(G)=P+1

Where P is the number of predicate nodes contained in the flow graph G.

c. Determine Independent Paths

The value of V(G) Provides the upper bound on the number of linearly independent paths through the program Control structural . Through the Control flow graph in Fig 5 we expect to specify six Paths:
Path 1: 1-2-10-11-13
Path 2: l-2-10-12-13
Path 3: 1-2-3-10-11-13
Path 4: 1-2-3-4-5-8-9-2
Path 5: 1-2-3-4-5-6-S-9-2
Path 6: 1-2-3-4-5-6-7-8-9-2

It is often worthwhile to Identify predicate nodes as an aid in the derivation of test cases. In this case, nodes 2,3,5,6, and 10 are predicate nodes [13].

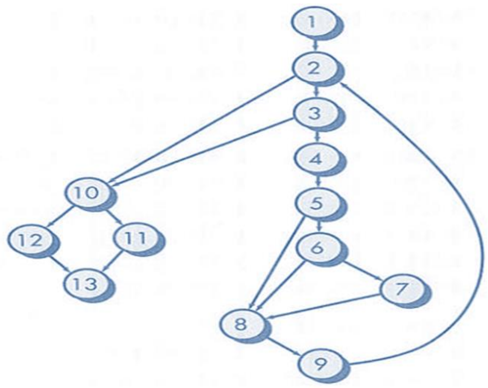


Fig 5 Control Flow Graph.

d. Deriving Test Cases

Data should be chosen so that conditions at the predicate nodes are appropriately set as each path is tested. Each test case is executed and compared to expected results. Once all test cases have been completed, the tester can be sure that all statements in the program have been executed at least once [13].

4. Intelligent Water Drop Algorithm

IWD algorithm [6][7] is a swarm-based optimization algorithm, simulated from observing natural water drops in river. IWD has been applied to various problems like Travelling Salesman Problem (TSP), N-queen puzzle, Multidimensional Knapsack Problem (MKP) , etc. These results have proved the significance of IWD algorithm over other swarm optimization algorithms. Another solution for TSP using IWD algorithm [7] is introduced where proposed algorithm converges very fast to the optimum solution.The improved IWD algorithm [4] has been applied to solve the air robot path planning in dynamic environments and results are quite impressive over genetic algorithm and ACO algorithm Since IWD has not yet been applied to the area of software testing and the effective results have been produced for various problems, this paper tries to derive a solution model for software testing using IWD in the hope that expected results will be more significant than the current solutions available for test data generation. Before moving to the proposed solution of IWD, general introduction is provided which describes its strategy along with available metrics in it.

IWD algorithm is a new swarm-based optimization algorithm inspired from natural rivers. In a natural river, water drops move towards center of the earth, due to some gravitational force acting on it. Due to this the water drop follows the straight and the shortest path to its destination [6].Pictorial representation of basic IWD is shown in Fig 6. In ideal conditions it is observed that the optimal path will be obtained. Water drop flowing in the river has some velocity which is affected by another actor, i.e., soil.

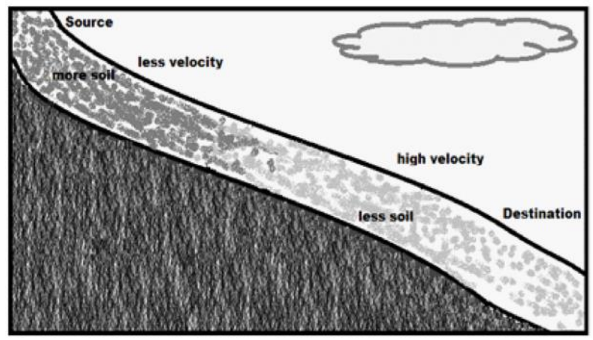


Fig 6 Pictorial representation of IWD.

Some changes that occurred while transition of water drop from one point to another point are:

1. Velocity of water drop is increased.
2. Soil content in the water drop is also increased.
3. Amount of soil in the riverbed from source to destination get decreased.

Water drop in the river picks up some soil in it when its velocity gets high and it releases the soil content when its velocity is less [7] Some of the prominent properties of the natural water drop are taken, based on which IWD is suggested. IWD has the two following important properties,

1. The amount of soil the water drop carries, which is represented by Soil (IWD) (or soil ^{IWD}).
2. The velocity of water drop with which it is moving now, denoted by Velocity (IWD) (or vel ^{IWD})

Value of both the properties may change during the transition. Environment contains lots of paths from source to destination [4] which may be known or unknown .When the destination is known, IWD follows the best path to reach the destination (best is in terms of cost and any other desired measure .(When destination is unknown it finds the

optimal destination. From the current location to the next location Velocity (IWD) is increased by an amount ‹which is nonlinearly proportional to the inverse of the amount of soil between the two locations ‹referred to as the change in velocity. The Soil (IWD), is also increased by extracting some soil of the path between two locations. The amount of soil added to the IWD is inversely (and nonlinearly) proportional to the time needed for the IWD to pass from its current location to next location. IWD chooses the path with less soil content. In the proposed approach, IWD is applied over the Control Flow Graph (CFG) to obtain the number of paths available in the program .The CFG depicts the logical control flow of the program [13]. All linearly independent paths could be obtained by CFG .Independent path is the path in the program that determines at least one new set of processing statement. In other words it introduces at least one new edge in the graph. Number of available paths can be obtained by finding the Cyclomatic complexity of the graph [13].
The IWD algorithm as specified by Shah-Hosseini H. in [6] is as follows:

1. Initialization of static parameters.
2. Initialization of dynamic parameters.
3. Spread the IWDs randomly on the nodes of the graph.
4. Update the visited node list of each IWD.
5. Repeat Steps a to d for those IWDs with partial solutions.
   a. For the IWD residing in node i, choose the next node j, which does not violate any constraints of the problem and is not in the visited node list of the IWD.
   b. For each IWD moving from node i to node j, Update its velocity.
   c. Compute the soil.
   d. Update the soil.
6. Find the iteration-best solution from all the solutions found by the IWDs.
7. Update the soils on the paths that form the current iteration best solution .
8. Update the total best solution by the current iteration best solution.
9. Increment the iteration number
10. Stops with the total best solution.

5. Ant colony Optimization Algorithm

The inspiring source of ACO is the food foraging behavior of real ants. When searching for food, ants initially explore the area surrounding their nest in a random manner. As soon as an ant finds a food source, it evaluates the quantity and the quality of the food and carries some of it back to the nest. During their return trip, ants deposit a chemical pheromone trail on the ground. The quantity of pheromone deposited, which may depend on the quantity and quality of the food, will guide other ants to the food source. The main principle behind these interactions is called stigmergy , or communication through the environment. An example is pheromone laying on trails followed by ants [3].
Pheromone is a potent form of hormone that can be sensed by ants while traveling along trails. It attracts ants and therefore ants tend to follow trails that have high pheromone concentrations. This causes an autocatalytic reaction, i.e., one that is accelerated by itself . Ants attracted by the pheromone will lie more of the same on the same trail, causing even more ants to be attracted see Fig7. This characteristic makes swarm intelligence very attractive for network routing, robotics, optimization etc. A number of extensions are proposed to the original ant algorithm. These algorithms performed better producing much improved results than the original ant algorithm [3].

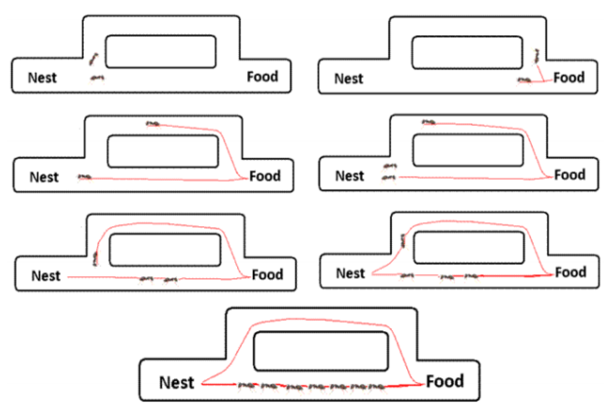


Fig 7 Optimization by Ant Colony.

Main characteristics of this model are positive feedback, distributed computation, and the use of a constructive greedy heuristic. The basic algorithm introduced by Marco Dorigo is given by following steps [3]:

1. Set parameters, initialize the pheromone trails
2. while (termination condition not met) do

a. Construct ants solutions
b. Apply local search
c. Update pheromones
d. end while

6. Proposed work

In the proposed work, it will be create a parser that used to convert input program to the corresponding control flow graph (CFG) for the program, then after knowing the benefits of intelligent water drop algorithm and their ability to find optimal solutions efficiently, it will be proposed to use it in the field of software testing through use in the generation of independent paths of the program and then, use ant colony optimization algorithm to generate the best test data that will be used in order to test all the independent paths the program and make sure it's have passed them all, as shown in Fig 8**.**

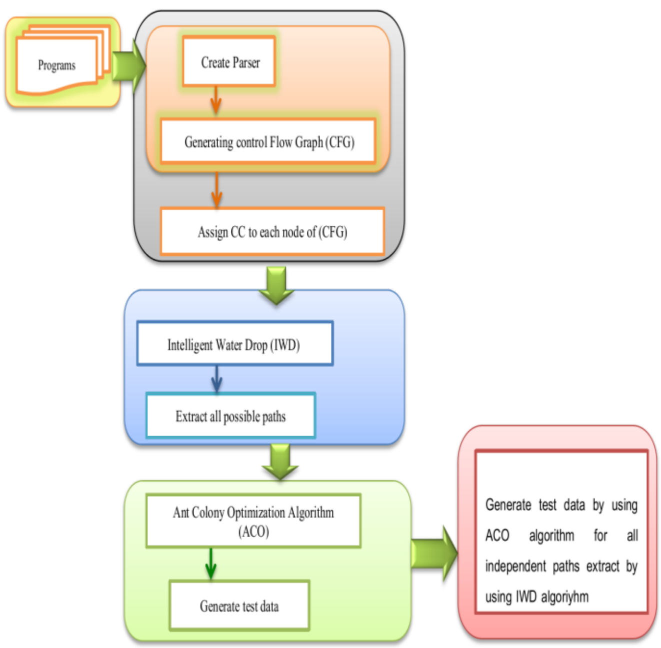


Fig. 8 proposed work

7. Conclusions

After a thorough study of swarm intelligence and its branches in particular IWD algorithm, know the benefits of IWD in the field of computer science and superiority over other techniques in this field, where it proved its ability by testing many of the real problems. So we proposed a method in the software testing process as they avoided the consumption of a large number of duplicates and their ability to reach solutions ideal and efficient manner as well as having knowledge of the importance of ACO algorithm and its ability to generate data that are used in software testing process, for all these reasons it will design a system that uses an Intelligent Water Drop Algorithm (IWD) to generate independent paths and test data in order to test all Independent paths in the program and make sure it's have passed them and covered efficiently.

References

[1] A. Windisch, S. Wappler, J. Wegener, "Applying Particle Swarm Optimization to Software Testing", ACM, 2007, pp. 1121-1128.
[2] F. N. Raza, , "Artificial Intelligence Techniques in Software Engineering"", In Proceedings of the International Multi Conference of Engineers and Computer Scientists, Vol.2174,2009, pp.1086-1088 .
[3] H. Afaq1 , S. Saini , " On the Solutions to the Travelling Salesman Problem using Nature Inspired Computing Techniques", IJCSI International Journal of Computer Science Issues, Vol. 8, No. 2,2011, pp. 326-334.
[4] H. Duan, S. Liu, X. Lei," Air Robot Path Planning Based on Intelligent Water Drops Optimization". IEEE International Joint Conference on Neural Networks , 2008, pp. 1397-1401.
[5] H. Schligloff , M. Roggenbach, " Path Testing" , Advanced Topics in Computer Science: Testing, citeseer ,2002.
[6] H. Sh. Hosseini ," The intelligent water drops algorithm: a nature-inspired swarm-based optimization algorithm", IJBIC International Journal of Bio-Inspired Computation, Vol. 1, No. 1/2, 2009,pp. 71-79.
[7] H. Sh. Hosseini, "Problem solving by intelligent water drops". IEEE Congress on Evolutionary Computation, 2007, pp. 3226-3231.
[8] H. Tahbildar ,B. Kalita , "AUTOMATED SOFTWARE TEST DATA GENERATION: DIRECTION OF RESEARCH", International Journal of Computer Science & Engineering Survey (IJCSES) Vol.2, No.1, 2011,pp. 99-120.
[9] Jovanovic , Irena," Software Testing Methods and Techniques",2008,pp.30-41.
[10] M. E. Khan, " Different Forms of Software Testing Techniques for Finding Errors", IJCSI International Journal of Computer Science Issues, Vol. 7, No 1, 2010, pp. 11-16.
[11] P. R. Srivastava , T. Kim," Application of Genetic Algorithm in Software Testing" ,International Journal of Software Engineering and Its Applications,Vol.3, No.4,2009,pp.87-96.
[12] R. S. Pressman , "Software Engineering A Practitioner's Approach, FIFTH EDITION",5th, McGraw-Hill Company,2001.
[13] R. S. Pressman , "Software Engineering A Practitioner's Approach, Seventh Edition",7th, McGraw-Hill Company ,2010.

[14] S. Nidhra, J. Dondeti," BLACK BOX AND WHITE BOX TESTING TECHNIQUES–A LITERATURE REVIEW" ,International Journal of Embedded Systems and Applications (IJESA) Vol.2, No.2, 2012,pp.29-50.
[15] S. S. Dahiya, J. K. Chhabra, Sh. Kumar , " Application of Artificial Bee Colony Algorithm to Software Testin21st Australian Software Engineering Conference, IEEE,2010,pp.149-154.
[16] S. Singla, D. Kumar, H M Rai, P. Singla," A Hybrid PSO Approach to Automate Test Data Generation for Data Flow Coverage with Dominance Concepts", International Journal of Advanced Science and Technology, Vol. 37, 2011,pp.15-26

First Author :**Dr. laheeb M. Alzubaidy**, have BSc. In 1987, MSc. In 1992 And PhD in 2002, in computer Sc. From Dept. of computer Sc, university of Mosul, Iraq. Associative professor in 2003, Head of Dept of Computer Sc. In 2003, visiting lecturer in Isra private university in 2004, head of Dept of Software Engineerinh in 2007, Visiting lecturer in USM university , NAV6 center in 2009, interested research fields are in Artificial Intelligent technique, network security , image processing , pattern recognition, software eng.

Second Author : Baraa S. Alhafid , have BSc. In 2006, in Software Eng.. From Dept. of Software Eng., university of Mosul, Iraq. Researcher in 2011, MSc. Student in 2011 Dept. of Software Eng., university of Mosul, interested research fields are in Artificial Intelligent technique, software eng.

ERP and Customization: Case Study of Logistics Processes Integration in a Telecommunications Company

Mustapha El Masbahi [1] and Alami Semma [2]

1 Department of Mathematics and Computer Science, Faculty of Science and Techniques, Hassan 1 University Settat, Morocco

2 Department of Mathematics and Computer Science, Faculty of Science and Techniques, Hassan 1 University Settat, Morocco

Abstract

The avoidance of customization is considered as critical success factor for ERP implementation. In this context, this paper aims to try to answer the following question: "ERPs are expected to adapt to business rules of the company in standard but why the use of customization is almost inevitable?" To address this question, we propose a case study of a "telecommunications" logistics processes integration in ERP. This study revealed an issue that is not supported in standard by ERP or mentioned in supply chain standards; then this paper propose a theoretical framework to formalize this issue.

Keywords: *ERP, implementation, customization, critical success factor, logistics, supply chain, telecommunications, information system*

1. Introduction

1.1. Context

The major challenge for the modern company is to optimize the supply chain: maximizing the use of resources and the customer service compatible with the business strategy. This is the reason for the continual recourse to the implementation of ERP in most companies [6]. Indeed, the implementation of an ERP carefully chosen can significantly reduce the cost of storage, the cost of raw materials, the customer delivery time, the lead time and the cost of production [7].

The fundamental principle of an ERP is to build computer applications in a modular way by sharing a single and common database. This creates an important difference with the previous situation (the customized applications existing before ERPs) because data is now assumed standardized and shared, which eliminates multiple entries and avoids the ambiguity of multiple similar data.

Note that ERPs cannot fit all companies; each of them is unique and has special management. Two solutions are possible:

- review the specific business rules to fit the ERP. This involves a preliminary analysis of processes and a reflection to be undertaken by the project owner and a good change management during the implementation of ERP. Indeed, employees must be prepared to see the nature of their work evolve, adopt new business rules and abandon the old software they used.
- use the customization by achieving specific developments, but don't go too far in specifics because in this case we lose the benefit of ERP. Indeed, the customization can be required only when the competitor advantage which involves the use of a non-standard process can be clearly justified [8].

On 1000 companies interviewed over customizing the ERP: 41% proceed to reengineering their business processes to fit the application, 37% choose the applications that fit their business processes and conduct minimal customization and only 5% customize the application to fit their business processes [9].

1.2. Motivation

Among the most active areas of research, we find that on the critical success factors for ERP implementation. The avoidance of customization is considered as critical success factor for ERP implementation [3]. The customization is not necessarily the parameterization but the use of specific developments.

Customizations usually generate an increase in the cost of information systems and the duration of the implementation and prevent us to benefit from upgrade and maintenance offered by the editor [4]. And that is why all companies who opt for the implementation of ERP seek at any price to avoid the specific developments but unfortunately this desire is still not fully answered, which leads us to ask the following fundamental question:

« ***ERPs are expected to adapt to business rules of the company in standard but why the use of customization is almost inevitable?*** »

To answer this question, we will work on a case study of logistics processes implementation in a telecom operator: this case study is a good example where the use of specific developments is inevitable.

1.3. Methodology

A considerable investigation has been conducted in a telecom operator over two years since the launch phase of ERP implementation project to the post-implementation phase. At the same time other auxiliary investigations were carried out in five different types of companies implementing ERP for different sectors. All these companies were examined to understand how they have implemented their logistics processes in the ERP.

After these studies, results have been found to help provide a theoretical framework as a basis to develop the models.

The paper is organized in two chapters: the first chapter presents the ERP literature and focuses on customization while the second chapter develops the issue raised by the case study.

2. ERP

2.1. Introduction

The information system of a company which does not have an ERP is typically composed of non-standard systems which are not found on the market. To support the development of the company, these systems should communicate with each other using interfaces that make after a while the information system of the company unoptimized.

In contrast to specific developments, ERPs have several advantages:

- optimization of business processes;
- coherence and consistency of information;
- integrity and uniqueness of the information system;
- sharing the same information system facilitates internal and external communication;
- minimization of costs: no interface between modules, synchronization of treatments, simplified corrective maintenance as provided directly by the editor and not by the IT department of the company;
- globalization of training;
- control of costs and deadlines for implementation and deployment.

The ERPs allow a company to manage and optimize all of its resources from production to after-sales service, through human resources, finance, logistics and sales. These complex projects lead to rethink existing organizations, to clearly identify the needs and to carefully plan the implementation of the information system.

2.2. Definitions

[36] For Deixonne, ERP is a software solution that provides the ability to manage all the resources of the company (human, material and financial) by focusing on two aspects: communication between different business actors and consistency of information. Similarly [39] Rosemann defines ERP as standard software solutions with integrated management for all processes within an organization namely production planning, warehouse management, finance, human resources management …

ERP is a generic and adaptable solution, developed by a single editor and consists of several interconnected modules integrating key functions needed to manage flows and procedures of the company (accounting and finance, logistics, business management ...).

The integration of business processes is considered one of the main objectives of the ERP implementation [26,27,28]. ERPs are indeed designed to address the problem of fragmentation of information in organizations [22]. A typical ERP integrates all functions of a business by allowing modules to share and transfer information freely [23,24]. ERPs do offer great opportunities to provide information to organizations in a standardized and centralized way [25].

Furthermore, the ERP is recognized as being effective in reducing inventory costs, improving efficiency and increasing profitability [29,26]. ERP is also known to reduce production time [30,31]. The ERP allows the stock declines, reductions in operating costs, the wealth of information desired by the customer and the ability to manage the extended company of suppliers, alliances and customers as an integrated whole [32]. With these advantages, the ERPs are widely implemented as backbone for many companies in the service and industry sectors [21].

2.3. Litterature

2.3.1. ERP Implementation and CSFs

[41] Botta-Genoulaz, Millet, Grabot (2000) classify the research literature on ERP systems in six categories: implementation of ERP, ERP optimization, management and ERP, ERP as a tool, ERP for supply chain management and case studies. They also noted that the post-implementation phase of projects, customization of ERP systems, the sociological aspects of the implementation, interoperability of the ERP with other systems and return on investment from the implementation are the most active areas of research.

The ERP system is a generic term for a large number of activities supported by multi-module software application that helps organizations to manage their resources [42]. However the ERP implementation is a complex exercise and many companies have encountered problems in the different phases [43,44]. In fact, many cases of failure in the ERP implementation due to cost and time overruns have been identified [45,46]. The high failure rate in implementing ERP calls a better understanding of the process [47]. In order to reduce the failure rate in implementing ERP, a number of studies have attempted to identify the critical success factors (CSFs) in ERP implementation.

[48] Bullen and Rockart defined CSF as "the limited number of areas in which satisfactory results will ensure competitive performance for the individual, department or organization. CSFs are the few key areas where things must function properly, so the business can thrive and the objectives of managers can be met". CSFs for ERP implementation bring a concept that helps an organization to identify critical issues that affect the process of implementation. Through a better understanding of the CSFs for ERP implementation, an organization can determine the corresponding solution to eliminate or avoid the causes of the most common failures in implementation [49].

For example, some CSFs identified in the case studies reviewed by Sumner [50] have included: the support of top management, the adaptation of business process to ERP standard, investment in training of the user, avoidance of customization and use of consultants with functional and technical expertise.

2.3.2. Customization

We propose in this section to examine customization of ERP systems which presents the area of research concerned by this paper. [52] According to Esteves and Pastor, the implementation phase concerns customization or parameterization and ERP package adaptation required to meet the needs of the organization. Often this task is performed with the help of consultants who provide implementation methodologies, know-how and training.

ERPs provide generic solutions to customers. They can more or less completely satisfy the needs of the company and especially when the business processes of the company are unique. And then customization is required. It is therefore important to choose the ERP that is easy to customize, so the time and cost consumed in the customization are minimized [58]. Customization is indeed a situation where a feature is added or modified not by the parameterization but by other means such as customizing the application code. In general, the modification of the ERP standard version in order to better support existing processes in an organization is referred to as customization.

Preserving the unique process is forcing companies to adopt ERP customization and take the risk of increasing the implementation time and the costs of maintenance and migration in the future [59]. However, companies must maintain a minimum customization since any change necessarily lead to higher costs [55] and more the software application is customized, more the cost and the possibility of error is high [56,57].

Generally customization increases the scope of the ERP project by adding time and cost to an implementation [54] and makes the software application unstable and difficult to maintain: a failure in the application can cause the cessation of the organization [53].

[60] Nah et al. (2001) list that minimizing customization is considered one of the CSFs for ERP implementation. And to minimize the need for customization, we must select at the acquisition phase the product that best meets the requirements of the organization [51].

New versions of the ERP may include new features that conflict with the changes already made or can remove structures that are required in the customized system of the organization [62]. This is why organizations implementing ERP should as far as possible try to adopt the features offered by the ERP rather than try to change the ERP to meet specific business practices [61]. CapGemini often advises that "it is cheaper to change the organization than to change SAP."

2.4. ERP implementation project

2.4.1. Context

The ERP implementation is a large project that undertakes for the enterprise very important installation and operating costs: licenses purchasing, acquisition of equipments, cost of implementation... Vision and objectives should be clearly stated in the business plan [66,67,83], including a justification of the investment, a clear statement of the project's mission and goals that should be associated with the business needs.

User habits can be conducted to change in this type of project in which a serious involvement of top management is strongly required to ensure achievement of objectives. For this, the top management must be willing to become involved and to allocate scarce resources to support the effort of implementation [66,67]. Indeed, ERP projects encroach on the boundaries between departments and affect many stakeholders, the top management was asked to mediate between different groups to solve political conflicts when necessary [68].

To be efficient and competitive, the company must consider the event of ERP implementation as a real opportunity to review its processes and to adopt good practices generally offered by ERPs.

2.4.2. Organization

The ERP implementation project cannot be ensured only by internal resources but necessarily requires the intervention of the integrator (supported by the editor) who provides the company functional and technical consultants bringing their expertise on ERP. The company must ensure the quality of consultants because it is considered as a critical success factor for the ERP implementation. As this project appeals to an external company, the organization of ERP implementation project is usually structured into two teams working in logic of customer/supplier relationship: project owner and project supervisor.

The project supervisor has the know-how and delivers ERP implementation; it then takes care of the training of key users, the animation of the needs analysis workshops and prototyping the adopted solution. While the project owner is responsible for monitoring the ERP implementation project and managing change which must be engaged early in the project. The quality assurance plan is a document often used to clearly identify the roles and responsibilities assigned to both project owner and project supervisor. It is recommended also that the project should be driven by the business department and not by the IT department [84,85,86].

2.4.3. Implementation

The ERP implementation is to configure and adapt the ERP to business processes identified and validated during the needs analysis with the various stakeholders and to put into production the finalized and accepted solution by key users. Parameterization and specific developments are obviously assured by the project supervisor and therefore a complete and clear documentation is essential: the documentation on parameterization must be performed throughout the project and in a very rigorous way, the used parameters, values and functional or technical meaning, interfaces with other systems, additions and modifications to the ERP must all be properly documented.

2.4.4. Change management

As long as the change management entails a real balance of power in favor of the change over the resisting forces [73], the company must be truly interested in changing policy to adopt at the beginning of the project. The company must indeed prove to users the choice of ERP and must carefully define why the ERP being implemented and what are the critical business needs addressed by the system [69,70,71,72].

Users generally adhere more to the ERP project than to the development project which generates naturally critics due to their technological and functional limitation compared to the ERP that offers more coverage and techniques from the latest technological advances.

A clear and effective communication at all levels of a company is required before and during the ERP implementation [82,67]. Communication includes the formal promotion of ERP project teams and the announcement of the progress of the project [66]. Amoako-Gyampah and Salam [81] note that effective communication is one of the success factors which act on the acceptance of the technology in an environment of ERP implementation.

The ERP requires the updating of skills that can only be achieved through training provided to employees of the company. If employees do not understand how a system works, they probably invent what they are capable of handling [74,75,76,77,78]. That is why the training is at the heart of change management and should be carefully managed. The company must consider both the initial training and recurrent training of users. Note also that user training will be effective only if it includes the business processes as a part of its content [79,80].

3. Case study

3.1. Introduction

During the investigation conducted in a telecom operator, we have examined the integration of logistics processes over all the phases of ERP implementation project.

In the phase of needs analysis, the logistics processes were carefully analyzed through several workshops with stakeholders: this phase helped to highlight the detail of data required for the logistics processes selected to be implemented eventually in the ERP. At this first phase, the integrator team warns that some features are not supported by the ERP standard, and then there were other workshops to prevent the use of specific developments but without success because the requested features are critical to the business and cannot be bypassed.

The use of specific developments was eventually chosen for the prototyping phase. The actual delay of this second phase was significantly higher than the expected delay because the specific developments were not anticipated previously.

After rolling the test phase, the project was put into production successfully. However requests for changes in the system have emerged during the post-production phase to ensure service continuity of the operator. We must recognize that the management of these changes has been difficult because of the rigidity of specific developments.

To understand why we had recourse to the specific developments, we first start by explaining the concept of “classical logistics” in the next paragraph, then we try in the paragraph that follows to describe the logistics processes of the operator while focusing on features not supported.

3.2. Description of « classical logistics »

Most companies, regardless of activity sector, proceed to purchase raw materials from suppliers to manufacture finished products for sale. Note that the raw materials, finished and semi-finished products are being stored in well-defined locations within warehouses. The implementation of this "classic logistics" is almost standardized by all ERPs.

The ERPs use the notion of item file to present the raw materials and finished products: an item must have a code that uniquely identifies it among other items codes handled by the company, be quantified according to a unit of measure and followed by a unique serial number.

After specifying the supplier for whom the purchase order of raw materials is intended, the purchase order transaction will include the item code to order, the ordered quantity and the unit of measure. When the delivery of supplier is performed, the receipt transaction must include the item code actually received, the quantity actually delivered by the supplier and the serial numbers associated with each unit if the item code needs to be followed by serial number.

The received items are sent to well-defined locations within warehouse: they can be moved or consumed on behalf of a work order for a finished product. Once the manufacturing process is complete, the finished product is stocked and ready to be delivered to customers. Likewise, the sale order transaction can materialize customer demand and includes information such item code, ordered quantity and requesting customer. For delivery to the customer, the shipping transaction must also specify the serial numbers actually shipped.

What we notice from the different transactions described above and that are implemented by all ERPs is that they all use the following information:

- Item Code
- Quantity
- Unit of measure
- Serial number (only invoked in the execution transactions namely receipt and shipping)

This set of information is a structural object supported by all supply chain modules of ERP.

3.3. Description of « telecom logistics »

The operations involved in logistics of a telecom operator are almost identical to those also found in classical logistics as it is described in the previous paragraph. Nevertheless logistics in a telecom operator has some features that deserve to be studied fine.

In fact, the items used in telecom logistics are mainly active items such as mobiles and smart cards for which phone numbers are associated. Logistically these items will have, identical to classical logistics, a code and a serial number and can be invoked in different transactions. These active items are characterized by technical attributes also called telecom attributes such as PIN, PUK, IMSI ... These attributes are useful for declaring active items on the network. This declaration is based on an information system which should draw its information from logistics operations. This suggests that the telecom logistics should convey telecom information (telecom attributes and phone numbers) and the structural set (item code and serial number) which is implemented by ERP standard.

In this context, telecom logistics requires the use of specific developments to be implemented in the ERP: the ERPs can only manage a structural set in standard.

Below are the main features of telecom logistics:

- before delivering the active items, the supplier first expects to receive telecom information from the operator, this information will be incorporated into the physical items;
- when the supplier performs the delivery, the active items come with other telecom information in addition to those provided by the operator;
- the operation of combining a phone number to the active item can be made by the supplier or by the operator;
- this operation can be related to an active item or a compound item including one or more active items, a compound item is indeed the result of a physical assembly operation.

We can naturally assume that the issue raised by the telecom logistics could also be encountered in other sectors for which the notion of item is not only limited to the physical component. The concept of item as it is modeled by ERPs is limited only to the structural set (code, serial number) that represents the physical component: the other piece of information we now describe as logical component receives only a passive storage (in the form of specific developments) and therefore it is not supported in standard by all modules.
According to another investigation on other ERPs, it also turns out that this feature is not supported in standard. Supply chain standards also don't address this issue: hence the need to generalize this issue.

3.4. Issue formalization

The objective of the supply chain management is to deliver the right product at the right time, in the right place and at a lower cost. This shows that the fate of the product is at the center of concerns for the supply chain management.
Note that the raw material is at the origin of the product and both are considered as objects clearly identifiable in the company.

The ERPs use the generic term "item" to model the raw materials, semi-finished and finished products: an item is indeed identified by a code. In addition, the ERPs permit to manage the stock which can be defined simply as all the items held by the company.

The inventory management aims to know at any time the items available in the company. To do so, it must provide physical accounting that takes into account the inputs and outputs of many items in order to provide, at any time, a reporting of the inventory update:

- Items input: this operation allows taking into account the input of items in the store: these received items can be manufactured (finished or semi-finished) or purchased (raw materials). The stock manager is responsible for updating the quantities of the items by increasing the initial quantity with the received quantity.
- Items output: this operation consists of removing the stock of items (requested by customers or production) in accordance with an order or an issue transaction. As for input, the stock manager should update the quantity of items by removing the output quantity of the initial quantity.
- Inventory reporting: at any time, the stock manager should be able to provide reporting of stocks. This reporting must show at a given time the detailed situation, quantity and location of stock.

The stock is the result of a difference between an input flow and an output flow over a period of time; we propose to formalize this statement as follows:

$$S_a(t_n) = S_a(t_{n-1}) + I_a[q] - O_a[q]$$

$S_a(t_n)$: is the stock (or quantity) of the item (a) at time (t_n)

$S_a(t_{n-1})$: is the stock (or quantity) of the item (a) at time (t_{n-1})

$I_a[q]$: is the item input of the quantity (q) of the item (a)

$O_a[q]$: is the item ouput of the quantity (q) of the item (a)

Note that this formalization above is supported in standard by all ERPs. The ERPs offer screens as items files to enter all required information. Other screens allow the entry of input/output transactions that affect the available quantities of stock in accordance with the above formula.

We cannot speak of the input/output operations without mentioning the other major production operation that is limited here as the equivalent of an input operation of the product and output operations of components. Indeed, the ERPs use the notion of bill of material that describes the composition of the product known as compound item. The composition of the product is the set of components items required to manufacture the product or compound item. As the previous statement, we propose to formalize this equivalence as follows:

$$M_{a_p}[q] \equiv \{I_{a_p}[q]\ ,\ O_{a_{c1}}[q]\ ,\ \ldots\ ,\ O_{a_{ci}}[q]\ \ ,\ \ldots\ ,\ O_{a_{cm}}[q]\}$$

$M_{a_p}[q]$: is production of the quantity (q) of the compound item (a_p)

$I_{a_p}[q]$: is the item input of the quantity (q) of the compound item (a_p)

$O_{a_{ci}}[q]$: is the item output of the quantity (q) of the component item (a_{ci})

We consider that the bill of material of the compound item (ap) mentions that there is (m) components items (aci) and for better clarity of the formula above it is assumed that the link quantity between compound and components is always equal to 1.

Don't forget that the issue of telecom logistics which leads to specific developments was informally described in a previous paragraph. And that is why we propose in this section to take this issue and try to re-express it formally with the help of the formalism above:

$$S_a(t_n) = S_a(t_{n-1}) + I_a \begin{pmatrix} inf(1,1) \ldots\ldots inf(1,k) \\ \ldots \\ inf(q,1) \ldots\ldots inf(q,k) \end{pmatrix} - O_a \begin{pmatrix} inf(1,1) \ldots\ldots inf(1,k) \\ \ldots \\ inf(1,1) \ldots\ldots inf(1,k) \end{pmatrix}$$

inf(i,j) represents the information required (known as telecom attribute) for item (a) and can be invoked in the input/output operations: each item (a) occurrence from (q) items has (k) information.

Indeed, the ERPs don't know how to take into account in standard the storage of this information in the same way as serial numbers which have an active storage that allows them to be raised in all transactions without resorting to even specific developments. Today's ERPs propose just a passive storage of this information with the help of laborious specific developments.

4. Conclusions

It is clear that today's companies are looking tirelessly for standard solutions sparing them the use of specific developments which don't support the scalability of information system and directly impact the business continuity.

Thanks to technological advances, ERP is now an essential tool for companies to be more flexible and responsive. And this is why editors, integrators constantly advocate that ERP presents a flexible and global solution and is able to respond to all business rules of the company in standard. In part this is true but there are still many cases exempt from this rule like the case study discussed in this paper. The case study has indeed revealed truths that deserve to be well highlighted in this conclusion.

Experience shows that all firms having implemented ERP are inevitably confronted with the exercise of choice between the re-engineering of business processes and the use of customization when it is impossible to support in standard one of the business rules. This exercise is even more difficult when the unsupported business rule is critical to the business. It is not denied that the re-engineering of business process according to a good practice proposed by ERP is the best solution when a business rule is irrelevant. In such situations, it is necessary to use the event of ERP implementation to adopt best practices.

When the business rule is relevant and critical for the company, the use of customization is inevitable, as is the case in our case study where the business rules are very specific to telecom sector. However, the specific developments have two major drawbacks: the costs of development and maintenance are very high and the editors don't propose any support. The ERPs should be often customized by functional actors rather than being modified by the technical actors.

Before drawing lessons from the issue raised by the case study, it is first important to note that the theoretical framework of classical logistics is largely covered by ERP; this was found after exploring the ERPs features and studying supply chain standards related [87]. In addition companies exploit only a small part of the features offered by ERP. It was also found that the best-known ERPs are based on the same model when implementing supply chain: in a perspective of widening the functional coverage of ERP, a minimal model is proposed in [87] and could serve as a good starting point to extend the model.

We can naturally assume that the issue raised by the telecom logistics could also be encountered in other sectors for which the notion of item is not only limited to the physical component. The concept of item as it is modeled by ERPs is limited only to the structural set (code, serial number) that represents the physical component: the other piece of information we now describe as logical component receives only a passive storage (in the form of specific developments) and therefore it is not supported in standard in all modules of ERP. It should also be noted that this issue is not addressed in the supply chain standards that are considered as a source of inspiration for ERP editors. The latter implement indeed the best practices as dictated by the standards in their software to fit to various sectors in standard and this therefore explains why all other ERPs don't take care of this issue.

To help theorists and practitioners, this issue has been formalized in a theoretical framework which is proposed in this paper and is supposed to be able to help generalize the issue and then evolve the standards.

In light of the foregoing, it can be deduced that the speed of standards and ERP updating is probably less than the speed of the market. This deduction could be considered as a possible answer to the fundamental question raised earlier in the introduction of this paper and reminded below:
"ERPs are expected to adapt to business rules of the company in standard but why the use of customization is almost inevitable?"

Similarly, the practices of some companies are not good and these companies refuse to admit this fact either by ignorance or for political considerations and they engage in specific developments; this could also be considered as other answer to the fundamental question.

The research community unanimously considers the minimization of customization as a critical success factor strongly affecting the ERP implementation, it is then highly recommended to enable the modern company to remain competitive by taking into account the specificities (resulting from the evolution of the market) which either are absent from standards or are not supported by ERP.

References

[1] Esteves-Sousa, J. and J. Pastor-Collado (2000). Towards the unification off critical success factors for ERP implementations. 10th Annual Business Information Technology conference, Manchester, UK.

[2] Moon, Y. b. (2007). "Enterprise Resource Planning (ERP): a review of the literature." International Journal of Management and Enterprise Development 4(3).

[3] M. Summer, « Critical Success Factors in Enterprise Wide Information Management Systems », Proceedings of the Americas Conference on Information Systems, Milwaukee, WI, 1999, pp 232-234

[4] M. A. Janson, and A. Subramanian, "Packaged Software: Selection and Implementation Policies", INFOR, 34, 2, 1996, pp. 133-151.

[5] S. Shankarnarayanan, ERP systems—using IT to gain a competitive advantage, March 23, 2000

[6] C. Loizos, "ERP: Is it the Ultimate Software Solution?", 1998, pp. 33-48.

[7] T. Gunn, "People: The Primary Resource in World Class Manufacturing," CIM Review, 4, 3, 1998, pp. 6-9.

[8] C. X. Escalle, and M. J. Cotteleer, Enterprise Resource Planning. Technology Note, HBS case #9-699-020, Harvard Business School, February 11, 1999.

[9] J. Davis, "Scooping up Vanilla ERP", Infoworld, 20, 47, November 23, 1998, 57+.

[10] Rosnay, J. d. (1975). Le macroscope : vers une vision globale. Points essais. Éditions du Seuil.

[11] Blumenthal, S. C. (1969). Management information systems : a framework for planning and development. Prentice-Hall.

[12] Le Moigne, J.-L. (1977). La théorie du système général. Presses Universitaires de France.

[13] Le Moigne, J.-L. (1992). La modélisation des systèmes complexes. Dunod.

[14] Mélèse, J. (1990). Approches systémiques des organisations. Éditions d'organisations.

[15] Shapiro Carl et Varian Hal, Information Rules, Harvard Business Press, 1998

[16] Shannon Claude E. - A mathematical theory of communication. Bell System Technical Journal (juillet-octobre 1948)

[17] Ackoff Russel, « Management Misinformation Systems », Management Science, 14, 1967

[18] Mélèse Jacques - L'analyse modulaire des systèmes de gestion. Hommes et Techniques (1972)

[19] Simon Herbert, The Sciences of the Artificial, MIT Press, 1969

[20] R. REIX, Systèmes d'information et Management des organisations, Vuibert, 1988

[21] Joseph R. Muscatello, Injazz J. Chen, Enterprise resource Planning (ERP) Implementations:theory and Practice, International Journal of Enterprise Information Systems, Volume 4, Issue 1 2008

[22] Muscatello, J. R., Small, M. H., & Chen, I. J. (2003). Implementing enterprise resource planning (ERP) systems in small and midsize manufacturing firms. International Journal of Operations and Production Management, 23(8), 850-871.

[23] Hicks, D. A., & Stecke, K. E. (1995). The ERP maze: Enterprise resource planning and other production and inventory control software. IIE Solutions, 27(8), 12-16.

[24] Chen, I. J. (2001). Planning for ERP systems: Analysis and future trend. Business Process Management Journal, 7(5), 374-386.

[25] Olson, D.L., Chae, B., & Sheu, C. (2005). Issues in multinational ERP implementations. International Journal of Services and Operations Management, 1(1), 7-21.

[26] Brakely, H.H. (1999). What makes ERP effective? Manufacturing Systems, 17(3), 120.

[27] Davenport, T.H. (1998). Putting the enterprise into the enterprise system. Harvard Business Review, July-August, 121-131.

[28] Davenport, T. H. (2000). Mission critical: Realizing the promise of enterprise systems. Harvard Business School Publishing.

[29] Appleton, E.L. (1997). How to survive ERP. Datamation, 43(3), 50-53.

[30] Goodpasture, V. (1995). Easton steps up to the plate. Manufacturing Systems, 13(9), 58-64.

[31] Davenport, T. H., & Brooks, J. (2004). Enterprise systems and the supply chain. Journal of Enterprise Information Management, 17(1), 8-19.

[32] Muscatello, J. R., Small, M. H., & Chen, I. J. (2003). Implementing enterprise resource planning (ERP) systems in small and midsize manufacturing firms. International Journal of Operations and Production Management, 23(8), 850-871.

[33] C. Ptak, E. Schragenheim, ERP: Tools, Techniques, and Applications for Integrating the Supply Chain, St. Lucie Press, Boca Raton, FL, 2000.

[34] S. Shankarnarayanan, ERP systems—using IT to gain a competitive advantage, March 23, 2000,

[35] H. Oden, G. Langenwalter, R. Lucier, Handbook of Material and Capacity Requirements Planning, McGrawHill, New York, 1993.

[36] Deixonne, J.-L. (2001). Piloter un projet ERP. Dunod.

[37] Tomas, J.-L. (1999). ERP et progiciels intégrés : La mutation des systèmes d'information. InterEditions.

[38] Lequeux, J.-L. (1999). Manager avec les ERP : Progiciels de gestion intégrés et Internet. Éditions d'organisation.

[39] Rosemann, M. (1999). ERP-software - characteristics and consequences 7th Europan Conference on Information Systems, Copenhagen, Denmark.

[40] Seddon, P. B., G. Shanks, et al., Eds. (2003). Second-Wave Enterprise Resource Planning Systems . Cambridge Cambridge University Press.

[41] V. Botta-Genoulaz, P.-A. Millet, B. Grabot, A survey on the recent research literature on ERP systems, Computers in Industry 56 (2005) 510–522

[42] A. Teltumbde, A framework for evaluating ERP projects, International Journal of Production Research 38 (2000) 4507–4520.

[43] M. Markus, S. Axline, D. Petrie, C. Tanis, Learning from adopters' experiences with ERP: problems encountered and success achieved, Journal of Information Technology 15 (2000) 245–265.

[44] Y. Xue, H. Liang, W. Boulton, C. Snyder, ERP implementation failures in China: case studies with implications for ERP vendors, International Journal of Production Economics 97 (2005) 279–295.

[45] V. Kumar, B. Maheshwari, U. Kumar, Enterprise resource planning systems adopting process: a survey of Canadian organizations, International Journal of Production Research 40 (3) (2002) 509–523.

[46] J. Scott, I. Vessey, Implementing enterprise resource planning systems: the role of learning from failure, Information Systems Frontier 2 (2000) 213– 232.

[47] T. Somers, K. Nelson, A. Ragowsky, Enterprise resource planning for the next millennium: development of an integrative framework and implications for research, in: Proceedings of the Americas Conference on Information Systems, Long Beach, CA, (2000), pp. 998–1004.

[48] C. Bullen, J. Rockart, A Primer on critical success factors, in: C. Bullen, J. Rockart (Eds.), The Rise of Managerial Computing: The Best of the Center for Information System Research, Dow Jones-Irwin, Homewood, Illinois, 1986, pp. 383–423.

[49] E.W.T. Ngai, C.C.H. Law, F.K.T. Wat, Examining the critical success factors in the adoption of enterprise resource planning, Computers in Industry (2008)

[50] M. Sumner, "Critical Success Factors in Enterprise Wide Information Management Systems", Proceedings of the Americas Conference on Information Systems, Milwaukee, WI, 1999, pp. 232-234.

[51]José Esteves, Joan Pastor, ENTERPRISE RESOURCE PLANNING SYSTEMS RESEARCH: AN ANNOTATED BIBLIOGRAPHY, Communications of the Association for Information Systems

[52] J. Esteves, J. Pastor, Enterprise resource planning systems research: an annotated bibliography, Communications of the Association for Information Systems, 2003, pp. 1–36.

[53] JD Thomson, Practical Studies In E-Governance: An Empirical Exploration Of Enterprise Resource Planning,International Review of Business Research Papers Vol.6, No.1 February 2010, Pp.432-466

[54] Toni M. Somers, Klara Nelson, The Impact of Critical Success Factors across the Stages of Enterprise Resource Planning Implementations, Proceedings of the 34th Hawaii International Conference on System Sciences - 2001

[55] Jiang Yingjie, Critical Success Factors in ERP Implementation in Finland, M.Sc. Thesis In Accounting The Swedish School of Economics and Business Administration 2005

[56] T. Davenport, Putting the enterprise into the enterprise systems, Harvard Business Review, July–August (1998) 121–131.

[57] F. Weston, ERP implementation and project management, Production and Inventory Management Journal 42 (3/4) (2001) 75–80.

[58] Zhe Zhang, Matthew K.O. Lee, Pei Huang, Liang Zhang, Xiaoyuan Huang, A framework of ERP systems implementation success in China: An empirical study, Int. J. Production Economics 98 (2005) 56–80

[59] B. Light, The maintenance implications of the customization of ERP software, J. Software Maintenance: Res. Practice 13 (2001) 415–429.

[60] Nah, F., Lau, J., & Kuang, J. (2001). Critical factors for successful implementation of enterprise systems. Business Process Management Journal, 7(3), 285-296.

[61] Parr A., Shanks G., Darke P. 1999. "Identification of Necessary Factors for successful Implementation of ERP

Systems". New information technologies in organizational processes - field studies and theoretical reflections on the future of work, Kluwer academic publishers, chapter 8, pp. 99-119.]

[62] Carlino J., Nelson S., Smith N. 2000. "AMR research predicts enterprise applications market will reach $78 billion by 2004", AMR research, Boston 2000.]

[63] Markus, M. L., Tanis, C. (2000). The Enterprise Systems experience - From adoption to success. In R. W. Zmud (Ed.), Framing the domains of IT management: Projecting the future through the past (pp. 173-207). Cincinnati, OH: Pinnaflex Educational Resources, Inc.

[64] Somers, T., Nelson, K., Ragowsky, A. (2000). Enterprise Resource Planning (ERP) for the next millennium: development of an integrative framework and implications for research. Paper presented at the 2000 Americas Conference on Information Systems, Long Beach, CA.

[65] Chang, S.-I., Gable, G. (2000). Major issues with SAP financials in Queensland government. Paper presented at the 2000 Americas Conference on Information Systems, Long Beach, CA

[66] C. Holland, B. Light, A critical success factors model for ERP implementation, IEEE Software 16 (1999) 30–36.

[67] F. Nah, K. Zuckweiler, J. Lau, ERP implementation: chief information officers' perceptions of critical success factors, International Journal of Human–Computer Interaction 16 (2003) 5–22.

[68] T. Davenport, Putting the enterprise into the enterprise systems, Harvard Business Review, July–August (1998) 121–131.

[69] J. Krupp, Transition to ERP implementation, APICS— The Performance Advantage (October)(1998)4–7.

[70] G. Latamore, Flexibility fuels the ERP evolution, APICS—The Performance Advantage (October)(1999)44–50.

[71] E. Schragenheim, When ERP worlds collide, APICS—The Performance Advantage (February)(2000)55–57.

[72] D. Travis, Selecting ERP, APICS—The Performance Advantage (June)(1999)37–39.

[73] P. Stebel, Breakpoints: How Managers Exploit Radical Change, Harvard Business School Press, Boston, 1992.

[74] Crucial success factors in an ERP makeover, Computerworld, November 29 (1999)45.

[75] H. Hutchins, 7 key elements of a successful implementation, and 8 mistakes you will make anyway, APICS 1998 International Conference Proceedings, Falls Church, VA, 1998, pp. 356–358.

[76] S. Laughlin, An ERP game plan, Journal of Business Strategy (January–February)(1999)32–37.

[77] C. Ptak, E. Schragenheim, ERP: Tools, Techniques, and Applications for Integrating the Supply Chain, St. Lucie Press, Boca Raton, FL, 2000.

[78] R. Sherrard, Enterprise resource planning is not for the unprepared, ERP World Proceedings, August, 1998

[79] T. Legare, The role of organizational factors in realizing ERP benefits, Information Systems Management Fall (2002) 21–42.

[80] D. Robey, J. Ross, M. Boudreau, Learning to implement enterprise systems: an exploratory study of the dialectics of change, Journal of Management Information Systems 19 (1) (2002) 17–46.

[81] K. Amoako-Gyampah, A. Salam, An extension of the technology acceptance model in an ERP implementation environment, Information & Management 41 (2004) 731–745.

[82] N. Bancroft, H. Selp, A. Sprengel, Implementing SAP R/3, Manning Publications Co., Greenwich, 1998.

[83] L. Zhang, M. Lee, Z. Zhang, P. Banerjee, Critical success factors of enterprise resource planning systems implementation success in China, in: Proceedings of the 36th Hawaii International Conference on System Sciences, 2003, pp. 236–245.

[84] W. Chew, D. Leonard-Barton, R. Bohn, Beating Murphy s law, Sloan Management Review (Spring)(1991)5–16.

[85] Crucial success factors in an ERP makeover, Computerworld, November 29 (1999)45.

[86] T. Minahan, Enterprise resource planning, Purchasing 16 (1998)112–117.

[87] Mustapha El Masbahi, Samir Bahsani and Alami Semma Enterprise Resource Planning (ERP) and Integration of Purchase, Inventory and Sale processes: Proposition of a Common Minimal Model IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 1, September 2012

[88] Abdelaali Himi, Mustapha El Masbahi, Samir Bahsani and Alami Semma A new approach to supply chain management based on pooling ITIL and APICS Principles and Practices IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 4, No 2, July2011

Fast Color Edge Detection Algorithm Based on Similarity Relation Matrix

H.I.ALI

Electronics and Communication Engineering Department, Helwan University, Cairo, 11792, Egypt

Abstract
A fast edge detection algorithm for color images was described in this paper. In the proposed method, smoothness of each pixel in color image is firstly calculated by means of similarity relation matrix and is normalized to maximum gray level. In spite of the simplicity of the proposed method, it can be further simplified taking benefit from the symmetry of matrix components as well as the avoidance of re-computing the common matrix elements between two successive pixel windows. This will reduce the time complexity extremely. The time complexity is further reduced by the splitting the algorithm into parallel tasks.
Keywords: *Edge detection; Similarity relation matrix; Smoothness thresholding*

1. Introduction

Edge is commonly defined as a sudden change in the local color intensity of an image or a jump in intensity from one pixel to the next. On the other hand, a region in image that generally represents objects could be defined as a collection of pixels, which share similar intensities. Thus boundaries of regions or objects in image are characterized by edges. The analysis of an image can be simplified by detecting edges since; it reduces the amount of data to be processed. Usually, edge detection is performed by smoothing, differentiating and thresholding.

The gradient of an image is the most common edge detector so far. Computing the gradient of an image can be performed by obtaining the partial derivatives in x and y directions by means of many operators like Roberts, Prewitt and Sobel operators [1, 2]. The gradient-based edge detection methods suffer from some practical limitations. Firstly, they need a smoothing operation to alleviate the effect of high spatial frequency in estimating the gradient. Smoothing is applied to all pixels in the image including the edge regions. This may distort or eliminate the edge in some cases. Secondly, the gradient magnitude alone is insufficient to determine meaningful edges because of the ambiguity caused by underlying pixel pattern, especially in complex natural scenes. Thirdly, the gradient-based edge detection method increases the computational complexity because calculations, such as square root and arctangent, to produce the gradient vector are required [1].

A detailed comparison and evaluation of edge detectors has been performed by Heath et al. [3]. They employed people to evaluate performance of several edge detectors with a number of images and looked for correlations in judgments of participants. Kim and Han have described edginess of pixels in terms of fuzzy rules [4] whereas the gradient magnitude and direction with fuzzy reasoning rules have been used to locate edges by others [5–7].

On the other hand, the edge detection process of color images is another important research issue. Typically, a color image consists of RGB channels. The color edge detection process must take into account the changes in intensity, chromaticity or both. So far, several color edge detection algorithms have been developed. These schemes can be classified into two different approaches. In the first approach, the three-channel image is processed as three gray-level images. We can use any gray-level edge detection scheme to detect the edge image for each color channel separately. Therefore, three edge images can be obtained for the RGB channels. Finally, a merging procedure is executed to combine these edge images into a targeted edge image. However, these sorts of algorithms have two major drawbacks. First, the inter-channel correlation is discarded in these schemes. Second, a high computational cost is consumed [8–10].

In the second approach, a two-stage structure is imposed on the design of color edge detection schemes. In the first stage, a channel reduction technique is employed to reduce the dimensionality of each color image from three to one. Next, an edge detection procedure for the reduced one-channel image is executed to detect edge. The two-stage edge detection schemes or color image have two advantages. First, the channel reduction process is independent of the edge detection scheme. The correlation among the color channels is taken into consideration. Second, any gray-level edge detection scheme can be applied to the detection of edge images in the second stage. In other words, the computational cost can be reduced [11–14].

Recep Demirci [15], proposed a two-stage edge detection algorithm. Firstly, the color image in three dimensional color spaces is mapped into one dimension by means of the similarity relation matrix. This transformation produces a gray level image where the similar pixels show the smooth areas and dark pixels show the dissimilar areas, noise and edges. Secondly, the thresholding could be employed if it is preferred. This work suffers from the high computational cost. So, in this paper we propose a fast edge detection algorithm that takes benefit from the symmetry of similarity matrix components as well as the avoidance of re-computing the common matrix elements between two successive pixel windows. Also a parallel implementation of the proposed algorithm is implemented which further reduces the time needed by the algorithm.

2. Color similarity

An image consists of pixels, which are neighbor to each other. Gray level differences of each color component between pixel P_1 and P_2 could be defined as follows:

$$\Delta R = |L_{R,1} - L_{R,2}| \quad (1)$$

$$\Delta G = |L_{G,1} - L_{G,2}| \quad (2)$$

$$\Delta B = |L_{B,1} - L_{B,2}| \quad (3)$$

The Euclidean distance can be computed as:

$$d_{i,j} = \frac{1}{\sqrt{3}} (\Delta R^2 + \Delta G^2 + \Delta B^2)^{1/2} \quad (4)$$

On the other hand, Wuerger et al. [16] showed in their research into proximity judgments in color space that perceptual color proximity is not Euclidean in nature. That means that distance information in Euclidean color space is not adequate for similarity judgment. The most general from of similarity measure based on the distance in color space could be given as below:

$$S_1(x_i, x_j) = 1 - \frac{d_{ij}}{D_n} = 1 - \frac{\|x_i - x_j\|}{D_n} \quad (5)$$

Where, S_1 is similarity between x_i and x_j, D_n is normalization coefficient. In Generalized Context Model, the similarity was expressed in terms of an exponential or Gaussian function of distance [17, 18]. So, the following formula could be obtained for color similarity:

$$S_2(x_i, x_j) = \exp\left(\frac{-d_{ij}^q}{D_n}\right) = \exp\left(\frac{-\|x_i - x_j\|^q}{D_n}\right) \quad (6)$$

With q=1 and 2 we obtain an exponential and a Gaussian functions respectively. The employment of Gaussian function with color distances to calculate similarity measure in term of color histograms was recently performed in applications for color image retrievals [28, 29]. Recep Demirci, showed that the usage of an exponential or Gaussian functions as color similarity functions, gives better performance in measuring similarity [15].

3. Proposed algorithm

3.1. Formation of similarity relation matrix

The similarity of any neighboring two pixels could be calculated by means of Eq.'s (5) and (6), respectively. A pixel in an image has eight neighboring pixels as shown in Fig. 1. Therefore, the similarity calculations for all the possible combinations are performed as shown in Fig. 2.

P1	P4	P7
P2	P5	P8
P3	P6	P9

Fig. 1 Pixel neighborhood

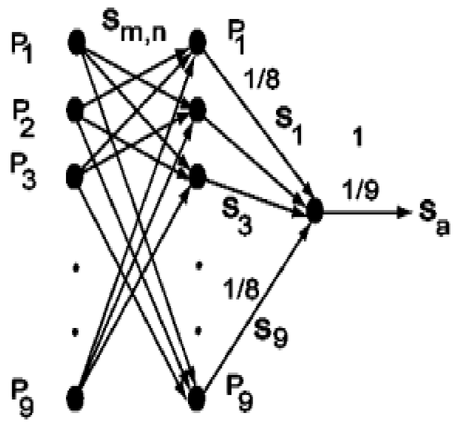


Fig. 2 Similarity network

This approach is well-suited with noisy exemplar approach proposed by Kahana and Sekular [17] where inter-stimulus similarity is used to categorize the noisy image. Consequently, similarity relation matrix is achieved as:

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 1, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

$$S_{m,n} = \begin{bmatrix} S_{1,1} & S_{1,2} & \cdots & S_{1,9} \\ S_{2,1} & S_{2,2} & \cdots & S_{2,9} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ S_{9,1} & S_{9,2} & \cdots & S_{2,9} \end{bmatrix} \qquad (7)$$

As the similarity of a pixel to itself is always equal to unity, there is no need to calculate them. The similarity relation matrix is symmetric. When all elements of the $S_{m,n}$, apart from diagonal, are unity, it means that the central pixel and its all neighbors have the same color level or it is perfectly smooth. The local smoothness of k^{th} pixel could be estimated as follows:

$$S_k = \frac{1}{8}\sum_{n=1}^{9} S_{k,n} \quad for\ k \neq n \qquad (8)$$

As could be seen from Eq. (8) and Fig. 2, there is no need to consider the similarity of itself. Therefore, Eq. (8) gives the average of how the k^{th} pixel is similar to the others. On the other hand, the general average of the central and all neighboring pixels could be calculated as follows:

$$S_a = \frac{1}{9}\sum_{k=1}^{9} S_k \qquad (9)$$

The Sa could be interpreted as the smoothness of the central pixel. Its values vary between zero and 1. On the other hand, the complement of the Sa (1−Sa) is considered to be dissimilarity or noisiness [15].

Calculating the similarity using this direct approach, as proposed in [15], needs very large time. So we proposed a modified approach for calculating the color similarity, which will reduce the computation time considerably. To show how to reduce the calculations needed to compute Sa, consider Fig. 3 that shows the elements of two successive pixel windows.

P1	P4	P7	P10
P2	P5	P8	P11
P3	P6	P9	P12

Fig.3 The elements of two successive pixel windows

If we need to compute the similarity of pixel P5, we will need to construct the similarity matrix shown in Eq. (7). Now to compute the similarity of next pixel P8, the similarities between (P4, P5, P6, P7, P8, and P9) need not be computed again. These similarities are already computed during the computation of the similarity matrix of pixel P5. So, we need to reshape this matrix to construct the similarity matrix for the next pixel (P8).

Studying the matrix in Eq. (7), we find that similarities between (P1, P2, P3) and all other pixels will not be needed again during the construction of the similarity matrix for the next pixel. So, the first step is to eliminate rows 1, 2 and 3, also the columns 1, 2 and 3. Secondly, the similarities between pixels (P10, P11, and P12) and (P4, P5, P6, P7, P8, P9) need to be computed.

Generally, the similarity matrix for pixel $P_{m+1,n}$ can be constructed as shown in Eq.(10). The lined elements will not be computed, so they will be cashed from the similarity matrix computed previously for pixel $P_{m,n}$. finally, we can summarize the steps to construct a similarity matrix for $P_{m+1,n}$ (call it $S_{m+1,n}$) using the similarity matrix for $P_{m,n}$ (call it $S_{m,n}$) as:

1- Eliminate rows 1,2,3
2- Eliminate columns 1,2,3
3- Compute the similarities between (P10,P11,P12) and (P4,P5,P6,P7,P8,P9)

$$S_{m+1,n} = \begin{bmatrix} \cancel{S_{4,4}} & \cancel{S_{4,5}} & \cancel{\cdots} & \cancel{S_{4,9}} & S_{4,10} & S_{4,11} & S_{4,12} \\ \cancel{S_{5,4}} & \cancel{S_{5,5}} & \cancel{\cdots} & \cancel{S_{5,9}} & S_{5,10} & S_{5,11} & S_{5,12} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cancel{S_{9,1}} & \cancel{S_{9,5}} & \cancel{\cdots} & \cancel{S_{9,9}} & S_{9,10} & S_{9,11} & S_{9,12} \\ S_{10,1} & S_{10,5} & \cdots & S_{10,9} & S_{10,10} & S_{10,11} & S_{10,12} \\ S_{11,1} & S_{11,5} & \cdots & S_{11,9} & S_{11,10} & S_{11,11} & S_{11,12} \\ S_{12,1} & S_{12,5} & \cdots & S_{12,9} & S_{12,10} & S_{12,11} & S_{12,12} \end{bmatrix} \qquad (10)$$

3.2 Local similarity thresholding (LST)

Kahana and Sekular [17] have described inter-stimulus dissimilarity as noise in their investigation by assuming that each stimulus is stored imperfectly in memory. Therefore they proposed to threshold the summed similarity of a stimulus to make decision. Keeping in mind that the complement of the Sa is considered as dissimilarity or noisiness and assuming that each pixel is stimulus, Recep Demirci [15], proposed a new version of similarity matrix as follows:

$$\tilde{S}_{m,n} = \begin{cases} 1 & if\ S_{m,n} \geq S_T \\ 0 & if\ S_{m,n} < S_T \end{cases} \qquad (11)$$

Where, S_T is similarity threshold. The interpretation of Eq. (11) is that if the similarity of two pixels is lower than S_T, they are considered to be dissimilar. So the local smoothness of k^{th} pixel could be estimated as follows:

$$\tilde{S}_k = \frac{1}{8}\sum_{n=1}^{9} S_{k,n} \quad for\ k \neq n \qquad (12)$$

The general average of the central and all neighboring pixels could be calculated as follows:

$$\tilde{S}_a = \frac{1}{9}\sum_{k=1}^{9}\tilde{S}_k \quad (13)$$

With such approach, a color image in three-dimensional color spaces is mapped into one-dimensional gray image while the noise is weakened and the edge information is preserved.

These calculations, as proposed in [15], could be reduced also, for the Euclidean, Exponential and Gaussian functions, with which we compute $S_{m,n}$. Our approach is based on rearranging the inequality (11) by putting the $\|x_i - x_j\|$ in one side, while all other variables are in the other side. In this way, we will obtain the following variations shown in table 1.

Table 1: Rearranging inequality (11) for Euclidean, Exponential and Gaussian forms

Original inequality [15]	$\tilde{S}_{m,n} = \begin{cases} 1 & if\ S_{m,n} \geq S_T \\ 0 & if\ S_{m,n} < S_T \end{cases}$
Rearranging the Euclidean form of $S_{m,n}$	$\tilde{S}_{m,n} = \begin{cases} 1 & if\ \|x_i - x_j\| \leq \mathrm{Dn}(1 - \mathrm{S_T}) \\ 0 & if\ \|x_i - x_j\| > Dn(1 - \mathrm{S_T}) \end{cases}$
Rearranging the exponential form of $S_{m,n}$	$\tilde{S}_{m,n} = \begin{cases} 1 & if\ \|x_i - x_j\| \leq -\mathrm{Dn} * \ln(\mathrm{S_T}) \\ 0 & if\ \|x_i - x_j\| > -\mathrm{Dn} * \ln(\mathrm{S_T}) \end{cases}$
Rearranging the Gaussian form of $S_{m,n}$	$\tilde{S}_{m,n} = \begin{cases} 1 & if\ \|x_i - x_j\| \leq \sqrt{-\mathrm{Dn} * \ln(\mathrm{S_T})} \\ 0 & if\ \|x_i - x_j\| > \sqrt{-\mathrm{Dn} * \ln(\mathrm{S_T})} \end{cases}$

Studying table1 carefully, we notice that the right hand sides of the inequalities are all of constant values. So, we can compute them once at the beginning of the algorithm and compare directly with the pixels difference. This will surely reduce the computation of $S_{m,n}$, since we need not to compute $S_{m,n}$, in its different forms, and then compare with ST. All what is needed is to compute the difference between pixels and directly compare this difference with the new constants, computed once, to get $\tilde{S}_{m,n}$.

4. Simulation results and discussion

The original algorithm [15] and the modified algorithm were tested with the well-known peppers image shown in Fig. 4(a) which is a color image of size 256 × 256 pixels. Fig. 4(b)–(d) show the smoothness image obtained by means of Eq. (5) with different normalization coefficients: 32, 64 and 128, respectively. Fig. 4(e)–(g) shows the transferred images when the similarity of pixels is calculated by means of exponential similarity function, Eq. (6), with different D_n: 32, 64 and 128, respectively. Moreover, Fig. 4(h)–(j) have been achieved when the Gaussian similarity measure is employed with different D_n: 1024, 8192 and 16 384, respectively.

Table 2 shows the time needed by the modified algorithm and the original algorithm [15] using same image, Peppers image. The algorithms were implemented on a computer having the following specifications: Intel(R) Core™2 Duo CPU @ 2,26GHZ. The results show that there is a large reduction in time complexity more than 40%.

Table 2: The time needed by the modified algorithm and the original algorithm

Distance function	Time for the original algorithm (seconds)	Time for the modified algorithm (seconds)	Reduction percent
Euclidean	155.54	92.32	40.65%
Exponential	158.25	93.31	41.04%
Gaussian	156.19	93.45	40.17%

Table 3 shows the time needed by the modified algorithm when implemented as parallel tasks on two processors and the original algorithm using same image, Peppers image. The results show that there is a large reduction in time complexity more than 70%. The image data was split into two equal parts. Each part is processed by a separate task on a separate processor.

Table 3: the time needed by the original algorithm and the modified algorithm when implemented as parallel tasks

Distance function	Time for the original algorithm (seconds)	Time for parallel implementation of the modified algorithm (seconds)	Reduction percent
Euclidean	155.54	42.64	72.59%
Exponential	158.25	43.95	72.23%
Gaussian	156.19	44.64	71.42%

Fig. 5 shows the transferred image of Peppers with original LST algorithm and the modified algorithm while normalization coefficient is fixed as 32. Fig. 5(a)–(c) show the images with ST: 0.25, 0.50 and 0.75, respectively, when linear similarity function is employed. Transformation of three dimension color space into one dimension with exponential function and different ST: 0.25, 0.50 and 0.75 have been shown in Fig. 5(d)–(f), respectively. Moreover, LST images obtained with Gaussian functions with ST: 0.25, 0.50 and 0.75 are shown in Fig. 5(g)–(i), respectively. It has been noticed from applications that the three dimensional color images could be successfully transferred into one-dimension image with these algorithms. The transformed image is a gray scale image in which the gray levels show the edginess strength.

Table 4 shows the time needed by the modified LST algorithm and the original algorithm using same image, Peppers image. The results show that there is a large reduction in time complexity around 40%.

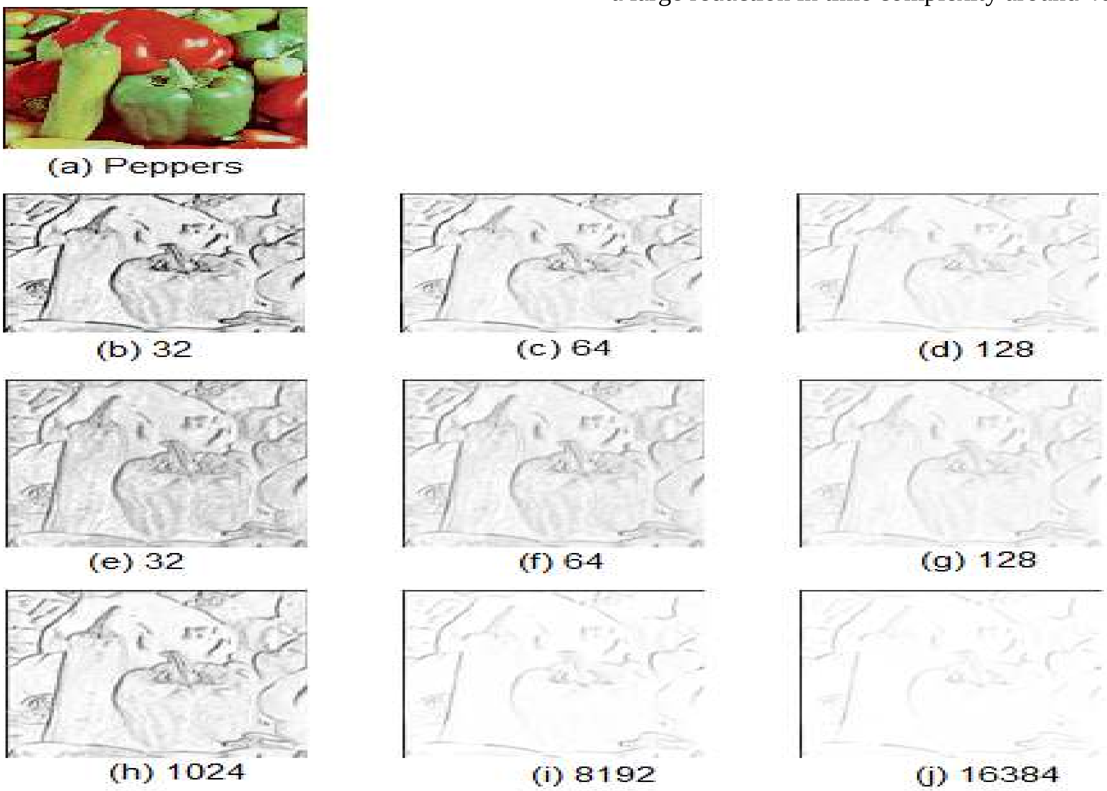


Fig. 4 Transformation of three-dimension color space into one dimension: Sa: (a) Peppers (b)–(d) By means of linear function with Dn: 32, 64 and 128. (e)–(g) By means of exponential function with Dn: 32, 64 and 128. (h)–(j) By means of Gaussian function with Dn: 1024,8192 and 16 384.

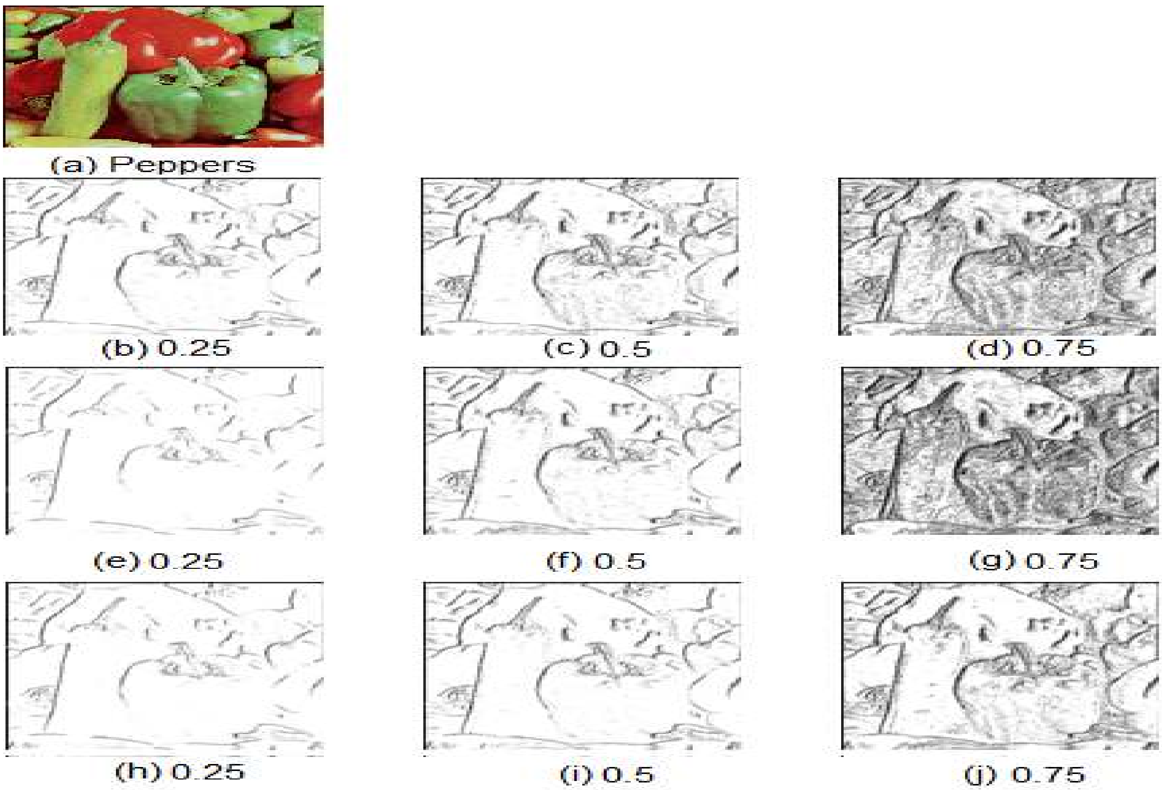


Fig. 5 Transformation of Lena image into one dimension with LST: Sa: (a) Peppers (b)–(d) By means of linear function with ST: 0.25, 0.50 and 0.75. (e)–(g) By means of exponential function with ST: 0.25, 0.50 and 0.75. (g)–(i) By means of Gaussian function with ST: 0.25, 0.50 and 0.75.

Table 4: The time needed by the modified LST algorithm and the original LST algorithm

Distance function	Time for the original LST algorithm (seconds)	Time for the modified LST algorithm (seconds)	Reduction percent
Euclidean	157.21	92.16	41.38%
Exponential	156.05	93.99	39.77%
Gaussian	157.30	93.84	40.34%

Table 5 shows the time needed by the modified LST algorithm when implemented as parallel tasks on two processors and the original algorithm using same image, Peppers image. The results show that there is a large reduction in time complexity more than 70%. The image data was split into two equal parts. Each part is processed by a separate task on a separate processor.

Table 5: the time needed by the original LST algorithm and the modified LST algorithm when implemented as parallel tasks

Distance function	Time for the original LST algorithm (seconds)	Time for parallel implementation of the modified LST algorithm (seconds)	Reduction percent
Euclidean	157.21	44.65	71.60%
Exponential	156.05	44.65	71.39%
Gaussian	157.30	43.26	72.50%

5. Conclusion

In this paper, a fast two-stage color edge detection algorithm, where the similarity relation matrix is used for the channel reduction process has been proposed. In spite of the simplicity of the original algorithm ,proposed in [15], it can be further simplified taking benefit from the symmetry of matrix components as well as the avoidance of re-computing the common matrix elements between two successive pixel windows. Accordingly, the three-dimensional color images could be successfully transferred into one-dimension images, which show color discontinuous as gray levels. The results show that there is a large reduction in time complexity more than 40%. The modified two-stage color edge detection algorithm was also implemented as parallel tasks on two processors. The results show that there is a large reduction in time complexity more than 70%.

References

[1] Gonzalez RC,Woods RE. Digital Image Processing. Reading, MA: Addison-Wesley; 1993.

[2] Canny JF. A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell 1986;8(6):679–98.

[3] Heath H, Sarkar S, Sanocki T, Bowyer KW. Edge detector comparison: initial study and methodology. Comput Vision Image Understanding 1998;69:38–54.

[4] Kim TY, Han JH. Edge representation with fuzzy sets in blurred images. Fuzzy Sets Systems 1998;100:77–87.

[5] Kim DS, Lee WH, S Kweon I. Automatic edge detection using 3×3 ideal binary pixel patterns and fuzzy-based edge thresholding. Pattern Recognition Lett 2004;25:101–6.

[6] Yuksel ME,Yildirim MT. A simple neuro-fuzzy edge detector for digital images corrupted by impulse noise. AEU—Int J Electron Commun 2004;58:72–5.

[7] Liang LR, Looney CG. Competitive fuzzy edge detection. Appl Soft Comput 2003;3:123–37.

[8] Fan J, Aref WG, Hacid MS, Elmagarmid AK. An improved automatic isotropic color edge detection technique. Pattern Recognition Lett 2001;22:1419–29.

[9] Ruzon MA, Tomasi C. Edge, junction, and corner detection using color distributions. IEEE Trans Pattern Anal Mach Intell 2001;23:1281–95.

[10] Theoharatos C, Economou G, Fotopoulos S. Color edge detection using the minimal spanning tree. Pattern Recognition 2005;38:603–6.

[11] Trahanias PE, Venetsanopoulos AN. Color edge detection using vector order statistics. IEEE Trans Image Process 1993;2:259–64.

[12] Yang CK, Tsai WH. Reduction of color space dimensionality by moment-preserving thresholding and its application for edge detection in color images. Pattern Recognition Lett 1996;17:481–90.

[13] Yang CK, Tsai WH. Color image compression using quantization, thresholding, and edge detection techniques all based on the moment-preserving principle. Pattern Recognition Lett 1998;19:205–15.

[14] Tsai P, Chang CC, Hu YC. An adaptive two-stage edge detection scheme for digital color images. Real-time Imag 2002;8(4):329–43.

[15] Recep Demirci, Similarity relation matrix-based color edge detection, Int. J. Electron. Commun. (AE-) 61 (2007) 469 – 477

[16] Wuerger SM, Maloney LT, Krauskopf J. Proximity judgments in color space: tests of a euclidean color geometry. Vision Res 1995;35:827–35.

[17] Kahana MJ, Sekuler R. Recognizing spatial patterns: a noisy exemplar approach. Vision Res 2002;42:2177–92.

[18] Stewart N, Brown GDA. Similarity and dissimilarity as evidence in perceptual categorization. J Math Psychol 2005;49:403–9.

Hossam Eldin Ibrahim received his B.Sc. degree in Communications & Electronics Engineering, his M.Sc. degree in Computer Engineering, and his Ph.D. degree in Computer Engineering from Helwan University, Cairo, Egypt, in 2000, 2004, and 2009 respectively. He has three published papers at M.Sc. degree, and seven published papers at Ph.D. degree. He is currently a Teacher at Electronics, Communication & Computer Department, Faculty of Engineering, Helwan University.

In-orbit Calibration Method Based on Empirical Model for Non-collinear TDI CCD Camera

Shujuan Zhang[1], Shuying Jin*[2]

[1] **Wuhan University Journals Press ,Wuhan University ,129 Luoyu Road,Wuhan 430079,China**

[2] **State Key Lab of Information Engineering in Surveying ,Mapping and Remote Sensing ,Wuhan University , 129 Luoyu Road,Wuhan 430079,China**

Abstract

An empirical model is proposed to solve the problem that exists in in-orbit calibration method for TDICCD camera in this paper. The method brings in the line frequency normalization and virtual image stitching from TDICCD images to the pre-processing procedure, which makes the virtual mosaic images has the same characteristics as the regular Linear push-broom CCD images, therefore the interior and external calibration is possible to realize. The experimental results show that the empirical model of TDICCD camera used in in-orbit calibration can improve precision and reduce the geometric distortion of images. Moreover, the geometric quality of images is enhanced. The method presented in this paper can be used to generate high-precision geometric rectification production.

Keywords: *TDICCD ; Internal Calibration ; External Calibration ; Empirical Model ; Geometric Accuracy*
.

1. Introduction

With the development of technique in high-resolution satellites, geometric positioning accuracy is much more concerned than ever. and is being one of the important assessment factor for satellites. According to the experiments, the main factors influencing the positioning accuracy of satellite images are showed as follow: （1） The precision of interior orientation ; （2） The precision of measured angle of camera， and (3)The precision of auxiliary data. Among which, the interior element affects geometric accuracy of the images, which leads to the position error presents in different sizes and orientations. And the overall geometric orientation precision are impacted by the remaining two factors. generally, camera’s interior elements describes internal calibration, and the Camera mounting angle describes exterior elements. The interior and exterior calibration can be accomplished in both laboratory or in-orbit with control fields. As the strong momentum coursed by Satellite launching and disturbances during satellite operation, the elements will change. So, it’s significant to use in-orbit calibration to guarantee satellite images with a stable geometric precision.

There are three types of internal calibration models for conventional push-broom linear array CCD cameras. Firstly, the physical parameters of the model has the trait that each coefficient has a clear physical meaning, while correlation between parameters lead to the normal equation morbid. Secondly, the self-calibration bundle adjustment with additional parameters. However, it’s difficult to establish an adaptive mathematical model for additional parameter entries and find out stable solutions. Thirdly, the linear array internal orientation models illustrated with high-order polynomial. All these models have a common process for inversing the control point coordinates to images based on collinear equation model. As the TDICCD(Time Delay Integration Charge Couple Devices) camera’s Line frequency changed over time, it’s complex to convert the coordinates based on the collinear equation model. What’s worse, iterations may not converge. Therefore, the current methods cann’t directly used for In-orbit calibration of TDICCD camera.

Aiming at the problems exist in in-orbit calibration of TDICCD camera, we propose a new empirical model based approach .This paper can be divided into five parts. Part 1 is the introduction , Part 2 reviews the strict imaging model of TDICCD cameras. The third part is the in-orbit TDICCD camera calibration method , Part 4 will show the results of experiments, and finally the conclusion and future work.

2.Strict Imaging Model of TDICCD Camera

Satellite CBERS1-02C was launched on December 22, 2011. It contains a high-resolution panchromatic camera with the resolution of 2.36m. Fig.1 shows the relative positions of three TDICCD in camera.

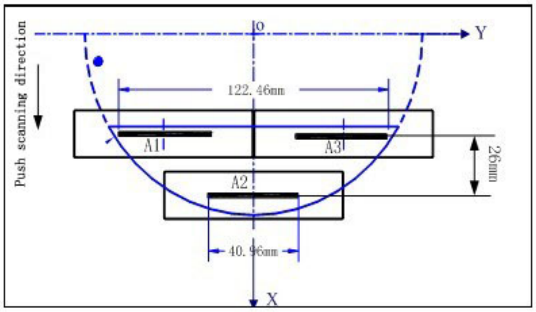


Fig.1 The three noncollinear TDICCD position on the focal plane of the camera

Compared with traditional CCD, the characteristics of TDICCD show as below:

1) It can adjust the integral series and line integration time (line frequency) in real-time, according to imaging areas and solar elevation angle at imaging time, to ensure the radiation quality of images.

2) During satellite operation, TDICCD can adjust the drift angle of the satellite platform in-time, to eliminate the motion blur of images. This makes the images have high clarity.

3) Shortcomings of TDICCD : the size of TDICCD is limited with the technology. In order to obtain larger image width, it can only be placed in non-collinear way. However, TDICCD placed in two rows cause the different image views of TDICCD and imaging time delay for the same surface features. These make geometric processing complicated. Fig.2 reflects this situation.

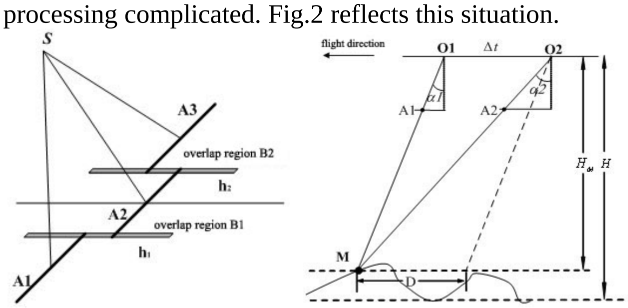


Fig.2 Imaging time delay with TDICCD placed in two rows of before and after

The stick imaging equation of each TDICCD has the following form.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{WGS84} = \begin{bmatrix} X_S \\ Y_S \\ Z_S \end{bmatrix}_{WGS84} + mR_T R_{GF} R_{FB} R_{BS} \begin{pmatrix} x_c \\ y_c \\ f \end{pmatrix} \quad (1)$$

Where, (x_C, y_C, f) are pixel coordinates in camera coordinate system, (X,Y,Z) are object coordinates of ground point in WGS84 geocentric cartesian coordinate system, (X_S,Y_S,Z_S) are object coordinates of satellite's center of mass in WGS84 geocentric cartesian coordinate system, m is the scale factor, R_{BS} is rotation matrix from camera coordinate system to satellite coordinate system with the origin at satellite center. R_{FB} is rotation matrix from satellite coordinate system to local orbital coordinate system. R_{GF} is rotation matrix from local orbital coordinate system to geocentric inertial cartesian coordinate system. R_T is rotation matrix from geocentric inertial cartesian coordinate system to WGS84 geocentric cartesian coordinate system.

And, R_{FB} and R_{GF} are determined by attitude and orbit measurement data. As attitude and orbit data is the temporal discrete points, it needs smoothing the data before processing. Then, fitting the data with polynomial or the Lagrange. Finally , calculating satellite's attitude and orbit at current imaging time with interpolating procession. The value of R_T is determined at imaging time and can be calculated by reference Standards of Fundamental Astronomy(SOFA).

3. In-orbit calibration method of TDICCD camera

3.1 Interior and external calibration

In the Eq.(1), the important factors which influence measurement precision of object coordinates obtained from images are auxiliary data error (such as like attitude error, the track error and time error), Camera installation angle error and the error of interior orientation. Within the camera's calibration is solving equations with using ground control points. In the process, the error from auxiliary data like R_{FB} , R_{GF} , R_T and (X_S,Y_S,Z_S) are processed as accidental error, at the same, the error from camera mounting angle and interior orientation that are contained in (x_C, y_C, f) and R_{BS} . External calibration is the process of calculating camera mounting angle or the rotate matrix R_{BS} , while interior calibration is to resolve elements of interior orientation. A strong correlation between elements of exterior and interior orientation of the various types of satellite imagery is found by studies. Thus, the problems like equation morbid may occur as the adjustment equations combined elements of interior and external orientation. Currently, a variety of solutions, like solving angle and line elements separately, merging strong-related items, ridge estimation method and using virtual observations to equations, are proposed for settling the problems of instable solution for normal equations as unknowns with strong dependencies.(Zhang Yongshen, etc., 2004) .In this paper, we adopt the method of solving parameters of exterior calibration and interior calibration stey by step.

3.2 Model of exterior calibration

Image geometric precision is consisted of the precision of interior orientation and exterior orientation. And, error of exterior orientation can be decomposed into the directions along the rail and vertical the rail. The positioning errors in the two directions corresponds to the error of camera mounting angle of φ_x, φ_y respectively. Though, the rotation matrix R_{BS} can be described with three angles, showed in Eq. (2), the error of exterior orientation will not be influenced by the rotation angle around the Z-axis. Therefore, External calibration only requires the solution of the error of φ_x, φ_y.

$$R_{BS} = R_{\varphi_x} R_{\varphi_y} R_{\varphi_z}$$
$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\varphi_x) & -\sin(\varphi_x) \\ 0 & \sin(\varphi_x) & \cos(\varphi_x) \end{bmatrix} \begin{bmatrix} \cos(\varphi_y) & 0 & -\sin(\varphi_y) \\ 0 & 1 & 0 \\ \sin(\varphi_y) & 0 & \cos(\varphi_y) \end{bmatrix} \begin{bmatrix} \cos(\varphi_z) & -\sin(\varphi_z) & 0 \\ \sin(\varphi_z) & \cos(\varphi_z) & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

Assuming N control points have been known, the model can be written as,

$$\begin{bmatrix} \bar{X} \\ \bar{Y} \\ \bar{Z} \end{bmatrix} = \frac{1}{m} R_{FB}{}^{-1} R_{GF}{}^{-1} R_T{}^{-1} \begin{bmatrix} X - Xs \\ Y - Ys \\ Z - Zs \end{bmatrix}$$

The Eq. (1) can be converted into the following form,

$$\begin{bmatrix} \bar{X} \\ \bar{Y} \\ \bar{Z} \end{bmatrix} = R_{BS} \begin{bmatrix} x_c \\ y_c \\ f \end{bmatrix} \quad (3)$$

Taking the partial derivatives of φ_x, φ_y. The error equation for control points can be written as .

$$\underset{3\times2}{A_i} \underset{2\times1}{Q} = \underset{3\times1}{L_i} \quad (4)$$

Where,

$$Q = \begin{bmatrix} d\varphi_x \\ d\varphi_y \end{bmatrix}, \quad A_i = \begin{bmatrix} \frac{\partial \bar{X}}{\partial \varphi_x} & \frac{\partial \bar{X}}{\partial \varphi_y} \\ \frac{\partial \bar{Y}}{\partial \varphi_x} & \frac{\partial \bar{Y}}{\partial \varphi_y} \\ \frac{\partial \bar{Z}}{\partial \varphi_x} & \frac{\partial \bar{Z}}{\partial \varphi_y} \end{bmatrix}$$

$$L_i = \begin{bmatrix} \bar{X} - \hat{\bar{X}} \\ \bar{Y} - \hat{\bar{Y}} \\ \bar{Z} - \hat{\bar{Z}} \end{bmatrix}, \quad \begin{bmatrix} \hat{\bar{X}} \\ \hat{\bar{Y}} \\ \hat{\bar{Z}} \end{bmatrix} = R_{BS} \begin{bmatrix} x_c \\ y_c \\ f \end{bmatrix}$$

,

The coefficient matrix of normal equation can be obtained from Eq.(5).

$$A^T A = \sum_{i=1}^{N} A_i^T A_i$$
$$A^T L = \sum_{i=1}^{N} A_i^T L_i \quad (5)$$

Solving the equations according to Eq. (6),

$$Q = (A^T A)^{-1} (A^T L) \quad (6)$$

$d\varphi_x, d\varphi_y$ are the corrections for φ_x, φ_y, and we get it from vector of Q.

Then, plusing $d\varphi_x, d\varphi_y$ to φ_x, φ_y. And, bringing the new value into list error equations, normalized and resolve equation, keeping on iterations until $d\varphi_x, d\varphi_y$ are less than given thresholds or iterative times reach a certain number .

3.3 In-orbit calibration method based on an empirical model for non-collinear TDI CCD camera

As known before, TDICCD in this satellite has the particularities of the change in line integral time and arranged non-collinear, we discuss two step preprocess of line frequency normalization processing and virtual image stitching for imageries at this situation. Line frequency normalization processing is that resampling the images as along rail direction. After it , each row of images has the same line integral time. That is to say, the difference of imaging line number for the corresponding point presented on three TDICCD imageries in the along-track direction is a constant. Consequently, in region of larger relief, the error of line number for the corresponding points is no more than 1 pixel. It's the base for stitching the imageries from the three TDICCD to virtual imagery. For mosaic images, the first step is to re-project the image from the center TDICCD to the imaging plane of the other two TDICCD. Thus, a complete line image is forming. Fig.3 shows describe the procedure.

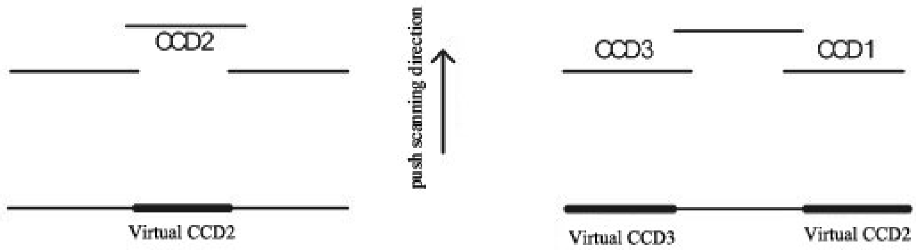


Fig.3 Virtual linear array image

The main point of this in-orbit calibration method is that it puts the two pretreatment methods into pre-processing process of the regular HR camera image. Thus, there exists no difference between the ordinary linear array push-broom CCD image and the virtual TDICCD image after image stitching. The virtual images can not only satisfy the imaging equation of Eq. (1) strictly, but also solving the exterior orientation elements with ground control points corresponding to the virtual image. What's more, rational polynomial model coefficients, which is equivalent to strict imaging model can be obtained with virtual linear array image , thus makes it possible to select ground control points using remote sensing image processing software (e.g. ERDAS) .

Assuming that there are N ground control points selected from the virtual linear array image by ERDAS software, and in accordance with part 3.2 of the external calibration, the formula (1) can be converted into the following forms:

$$\begin{bmatrix} x_c \\ y_c \\ f \end{bmatrix} = \frac{1}{m}(R_T R_{GF} R_{FB} R_{BS})^{-1} \begin{bmatrix} X - X_s \\ Y - Y_s \\ Z - Z_s \end{bmatrix} \qquad (7)$$

The internal calibration model for the normalized cubic polynomial is showed as following.

$$\begin{aligned} \frac{x_c}{f} &= a_0 + a_1 \times s + a_2 \times s^2 + a_3 \times s^3 \\ \frac{y_c}{f} &= b_0 + b_1 \times s + b_2 \times s^2 + b_3 \times s^3 \end{aligned} \qquad (8)$$

Where,

S is the number of unit on the virtual image; $\frac{x_c}{f}, \frac{y_c}{f}$ are the calculated from Eq.(7) according to control points. $a_0, a_1, a_2, a_3, b_0, b_1, b_2, b_3$ are coefficients for interior alibration, which are unknowns and composed with translation of sensor linear array relative to the camera main point, the size of CCD units, the tilt of CCD lines and lens' non-linear distortion. Then, the directions of each CCD unit in camera coordination can be computed in account of these coefficients. The initial value of the coefficients is given according to the design of the camera's factory or laboratory calibration results.

It's remarkable that, $\begin{bmatrix} x_c \\ y_c \\ f \end{bmatrix}$ presented the direction in a three dimension space in Eq. (1) is a vector, and the length of it is not a constant. $\begin{bmatrix} x_c \\ y_c \\ f \end{bmatrix}$ required by the ray of light intersecting on WGS84 reference ellipsoid. It needs to transform $\begin{bmatrix} x_c \\ y_c \\ f \end{bmatrix}$ to $\begin{bmatrix} \frac{x_c}{f} & \frac{y_c}{f} & 1 \end{bmatrix}^T$ when submitting $\frac{x_c}{f}, \frac{y_c}{f}$ obtained from interior calibration into Eq. (1).

4. Experiments

The experimental images of this article is from HR camera on CBERS1-02C Satellite, and reference data of field control points, high-precision aerial ortho photo and DEM are from the ground calibration field in Songshan, China.

It shows the distribution of experimental imageries and control points in Fig.4.

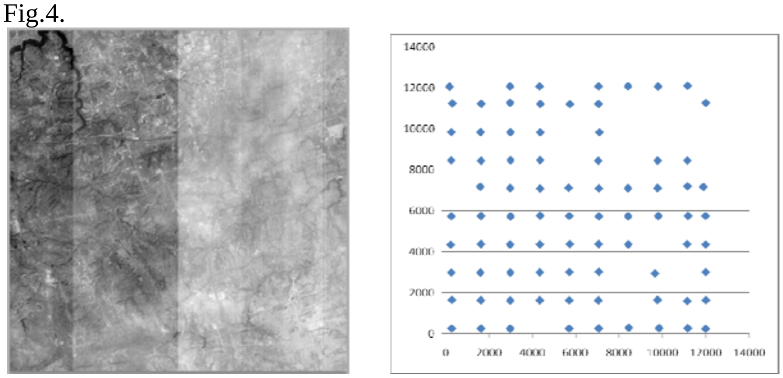
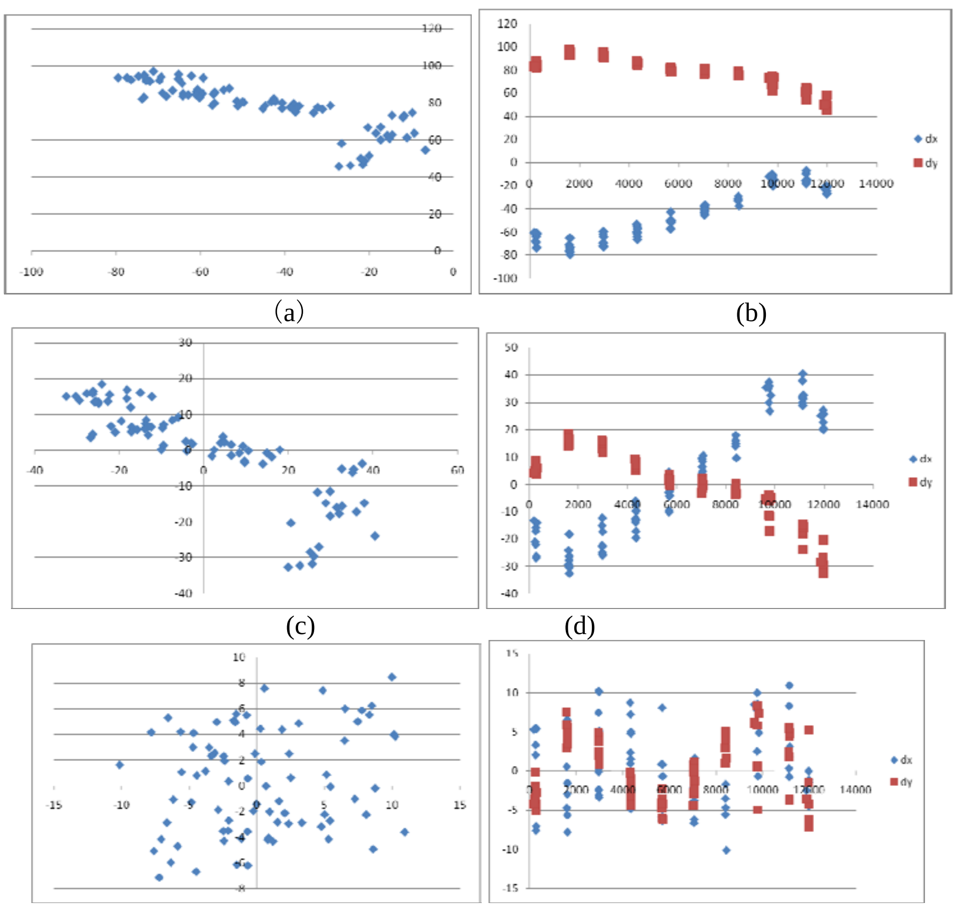


Fig. 4 The distribution of experimental imageries and control points.

Experimental steps are as follows.
1) Normalizing line frequency of images.
2) Stitching images to form a whole virtual image.
3) Choosing field control points and calculating the error of them.
4) Recomputing the error of field control points after external calibration.
5) Recomputing the error of field control points after interior calibration.
Fig.5 shows the results.



(a) (b)

(c) (d)

(e) (f)

Fig.5 (a), (c), (e) present the error distribution of control points. with horizontal axis representing the error perpendicular to the rail direction and vertical axis for error along the rail direction. Figure (b), (d),(f) shows trends of the error of control points in the two directions as the change of the virtual image row number. The error presented in figure (a) and (b) is calculated without calibration process. The error displayed in figure (c), (d) is obtained after external calibration. And the error shown in (e) ,(f) is required after both external and interior calibration.

Table1: The change of control point error before and after the scaling process (unit:pixel)

		Vertical the rail direction	Along the rail direction
Initial error	Mean	-47	+79
	Standard deviation	22	13
Error after external calibration	Mean	0	0
	Standard deviation	22	13
Error after external and interior calibration	Mean	0	0
	Standard deviation	5	4

5. Conclusion

In consideration of image feature of satellite CBERS1-02C, this article proposed a new calibration method based on non-linear TDICCD with an empirical model. The distinguishing feature of this method is to combine the line frequency normalization to virtual images stitching from three TDICCD images in pre-processing step. Thus, the mosaic image has the same characteristics with the common imageries from linear push-broom CCD It's the key of this article to realize interior and external calibration.

According to the experiment, this in-orbit calibration method based on an empirical model for non-collinear TDI CCD camera can not only improve the precision of exterior orientation, but also lower the geometric distortion of images. It's possible to generate high-resolution geometric rectification production directly with this method.

However, as the reasons that the precision of satellite attitude measurement is not high and the intensity of attitude measurement point is not dense enough, resulting in the drift angle error of the satellite platform and leads to the number of overlapping pixels from two adjacent TDICCD images changes. and the elements of interior orientation of the virtual image are unstable.

How to obtain the drift angle by the number of overlapping pixels of two adjacent TDICCD images accurately, is the problem we need to solve in further study.

Acknowledgements

The Project Supported by the National Natural Science Foundation of China (Grant No.40901209).

References

[1]J. Grodecki, G Dial, J Lutes. Mathematical model for 3D feature extraction from multiple satellite images described by RPCs. ASPRS Annual Conference Proceedings,2004

[2]J.Grodecki.IKONOS stereo feature extraction-RPC approach.Proc. ASPRS Annual Conference, St. Louis, 2001

[3]C.Yu,M.L.J Apuzzo,C.S Zee,Z Petrovich.Instrumentation Assessments-A Phantom Study Of The Geometric Accuracy of Computed Tomographic and Magnetic Resonance Imaging Stereotactic. Neurosurgery,2001

[4]C .V. Tao, Y. Hu. 3D reconstruction methods based on the rational function model. Photogrammetric Engineering and Remote Sensing, 2002

[5]C. V. Tao, Y Hu .The rational function model-A tool for processing high-resolution imagery. Earth Observation Magazine, 2001

[6]J． Höhle. Photogrammetric measurements in oblique aerial images. Photogrammetrie, Fernerkundung, Geoinformation, 2008

[7]G. Dial, J. Grodecki .IKONOS accuracy without ground control. Proceedings of ASPRS Comission1 mid-term,2002

[8]Y. Oh, K. Sarabandi, F.T Ulaby .An empirical model and an inversion technique for radar scattering from bare soil surfaces. Geoscience and Remote,1992

First Author: Shujuan Zhang Received her BE degree in photogrammetry and remote sensing from School of Remote Sensing Engineering ,Wuhan University and her ME degree from school of Resources and Environmental Science ,Wuhan University. She is currently an associate professor in Wuhan University.She has published more than 15 papers.Her research interests include photogrammetry and remote sensing.

Second Author: Shuying Jin received her Ph.D. in photogrammetry and remote sensing from Wuhan University.She is currently an associate professor in Wuhan University.Her studies have been focused on photogrammetry and remote sensing.

A 3D ear recognition method based on auricle structural feature

Kai Wang, Zhi-chun Mu

School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, 100083, China

Abstract

The performances of most existing 3D ear recognition methods are degraded sharply by pose variation. In this paper, a 3D ear representation called 3D auricle structural feature(3DASF) and the corresponding pose robust 3D ear recognition method is presented. By measuring the surface characteristics through Surface Variation, 3DASF that contains ear key physiological structure is extracted. Then 3DASF corresponding points are used to implement iterative closest point(ICP) algorithm to coarse align gallery-probe ear pairs. Finally, fine alignment is performed to obtain the alignment errors for identity recognition. Experimental results conducted on University of Notre Dame(UND) biometric datasets collection F and collection G outperform the state-of-the-art 3D ear recognitions based on ICP. The results also demonstrate that the proposed method is more robust to pose variation than the state-of-the-art.

Keywords: *3D ear recognition, 3D feature extraction, auricle structure, iterative closest point.*

1. Introduction

Biometric system uses human physiological characteristics, such as face[1], fingerprint, iris etc., as well as behavioral characteristics, such as gait, handwriting, voice authentication etc., which can not easily be stolen and counterfeited. As a biometric trait, the human ear, has the common characteristics of biometric traits, such as universality, uniqueness, permanence etc. , has a rich and stable structure, and unlike the face which is variable under different scenario such as expression variation, aging, makeup and eyeglasses, is receiving more and more attentions. Ear recognition based on 2D images in constrained condition has attained the accuracy of 90% , however, 2D images of the human ear is susceptible to variation in illumination and pose, greatly affecting the recognition accuracy. The research of Chang and Bowyer show that ear recognition based on 2D images is greatly affected by pose variation, even reduced to below a accuracy of 30%[2]. How to improve the accuracy in the case of pose variation is one of hot issues in theory and application of ear recognition.

Using the commercial 3D laser scanner to obtain the 3D shape of human ear for recognition, which exploits the depth information, also is propitious to overcome the problems such as illumination and pose variation. In the literatures about 3D ear recognition, ICP[3] is most widely used, that is a classical method of 3D point cloud registration. Thus, the most of early research work are testing and evaluating the different variants of ICP applied to 3D ear recognition[4-9]. Recently, the framework of 3D ear recognition utilizing 3D feature extraction has gained much attentions. Bir Bhanu et al. put forward the LSP (Local Surface Patch) descriptors[10], which has shown good results in free form object classification, and successfully applied to 3D ear recognition[11-13]. Islam et al. employ local 3D features to achieve coarse alignment of 3D ears. Moreover, a large number of non candidate samples can be excluded according to the alignment errors[14]. The aforementioned researchers both use the extracted local surface features to obtain the ear corresponding points, but a satisfactory recognition accuracy is difficult to be obtained when only the corresponding points between ear samples are used according to the experiment results. In order to make good use of the acquired corresponding points, the coarse alignment of point cloud is performed with these points, then fine alignment is done with ICP algorithm, thus good recognition accuracy can be achieved. Zhou et al.[15] propose the Surface Patch Histogram of Indexed Shapes (SPHIS) 3D feature descriptors for local ear surface representation and matching, then voxelization is used to holistic matching, finally the scores from local and holistic match components are fused to generate the final scores. Comparing to the Hui[11] and Islam[14], Zhou et al.‘ method achieves better results.

The methods of 3D ear recognition based on 3D feature extraction improve the computational efficiency, however, good performances only can be achieved in datasets with pose nearly invariability. In the occasion of pose variation, those methods perform not well. Aiming to this problem, we propose a new 3D ear feature for recognition-- 3D auricle structural feature(3DASF) and the corresponding

3D ear recognition method. The coarse alignment is efficiently carried out with the 3DASF points, then ICP fine alignment is performed to obtain the alignment error. The experiments on University of Notre Dame(UND) biometric datasets collection F and collection G are performed to evaluate the proposed method. Comparisons with state-of-the-art methods are performed to illustrate the applicability of the proposed method.

The rest of paper is organized as follows. The 3D auricle structural feature is described in Section 2. The 3D ear recognition based on auricle structural feature is explained in Section 3. The performance of the proposed method is evaluated and compared with other approaches in Section 4. Finally, Section 5 gives the conclusions.

2. 3D Auricle Structural Feature

The performance of a biometric recognition system directly depends on how to represent the data and how to extract the important feature from the representation of data. Ear has abundant ridges and channels which are shown as the concave and convex surface. Meanwhile, the key physical structure of ear, such as helix, antihelix, earlobe, etc. mostly are located at the vertices of the convex or concave surface, which can be seen from Figure 1. The different colors represent the variation of ear surface depth. If the depth variation can be estimated, based on this, the surface vertices can be classified into different types, and the key physical structure could be extracted.

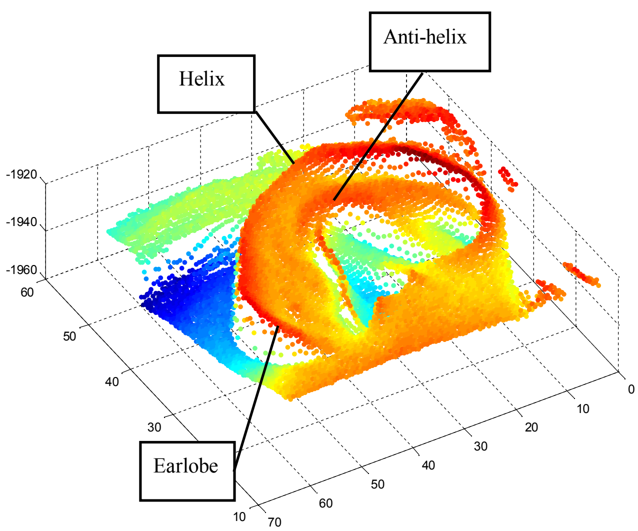


Fig. 1 3D auricle structure

2.1 Representation of 3D ear and local neighborhood of vertex

Generally, 3D data is represented and storaged by point cloud or mesh. The point cloud produced by 3D laser scanner can be used to construct mesh to obtain more compact storage and representation. The experimental data used in this paper comes from the University of Notre Dame, UND biometrics dataset, collection F and collection G[16], which was collected using Minolta Vivid 910 laser scanner. All the data in UND dataset are single scanning point clouds. The single scanning point cloud of ear from a singe view usually has self occlusion, nevertheless, the parts that are prone to be self occluded often are located at cavum conchae and the concave surface between helix and antihelix, that will not have a crucial impact to the completeness of the key physical structure of ear. Therefore, we first uniformly resample the raw point cloud, and use grid fitcode from mathworks website[17] to conduct surface fitting to transform point cloud into depth lattice array. Then mesh representation of 3D ear which has not geometrical sudden changes and still be consistent with the 3D ear shape ground truth is obtained by triangulation, as shown in Figure 2.

In order to estimate the variation of depth, the neighborhood of 3D mesh vertices is defined as follow. For a mesh vertices v_0, a set of rings can be defined around it, the first ring R_1 includes all immediate neighbors of the vertices v_0, the second ring R_2 including the all immediate neighbors of the vertices in the first ring c and so on. The c-th ring R_i can be defined in the following condition: for one vertex v of i-th ring c, there exists a shortest path from v_0 to c that contain i

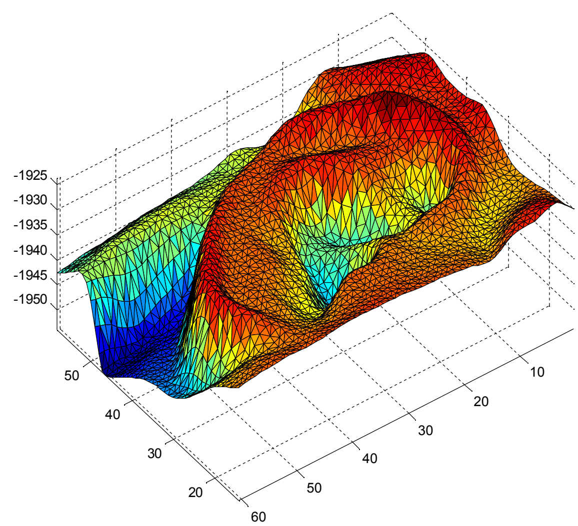


Fig. 2 The mesh representation of a 3D ear

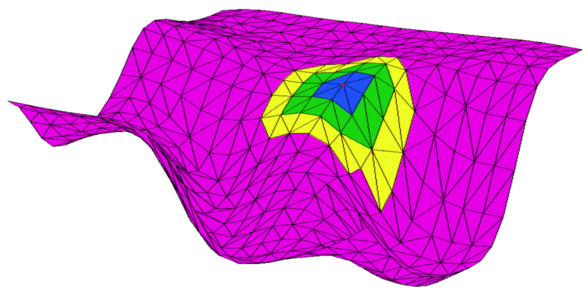


Fig. 3 3-ring neighborhood. The meshes in 3 rings are shown with different colors.

edges, then the set $\Re = \{R_i : i \le N\}$ is defined as the N-ring on the vertices v_0. Figure 3 shows a 3-ring neighborhood around a mesh vertex.

2.2 Surface Variation

Eigenvalues of the covariance matrix of a local surface neighborhood can be used to estimate the surface properties[18, 19]. A vertex of 3D mesh and its N-ring neighborhood vertices is given by $p_i = [x_i \; y_i \; z_i]^T$ ($i = 1, \dots, n_l$), then the covariance matrix is given by:

$$C = \begin{bmatrix} p_{i_1} - \overline{p} \\ \cdots \\ p_{i_k} - \overline{p} \end{bmatrix}^T \cdot \begin{bmatrix} p_{i_1} - \overline{p} \\ \cdots \\ p_{i_k} - \overline{p} \end{bmatrix}, \quad i_j \in N_p \qquad (1)$$

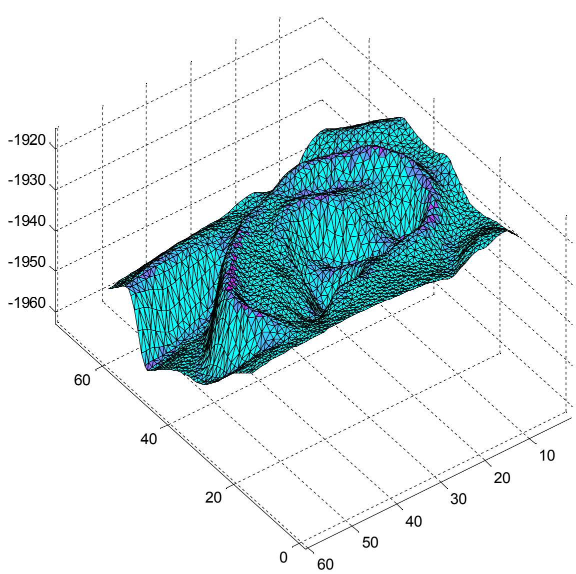


Fig. 4 3D auricle structural feature base on surface variation, as shown in purple.

Where $\overline{p}$ is centroid of the neighbors p_{i_j}, eigenvectors v_l can be obtained through eigenvalue decomposition of covariance matrix C by Principal Component Analysis(PCA). $C \cdot v_l = \lambda_l \cdot v_l, l \in \{0,1,2\}$.

C is a symmetric and positive semi-definite matrix, all the eigenvalues λ_l are real numbers, eigenvectors v_l consist an orthogonal basis corresponding to the primary components of vertices set N_p. v_l measure the variation of p_i, $i \in N_p$ along the direction of corresponding eigenvectors. The total variation is defined as sum of the square distance between p_i and its centroid:

$$\sum_{i \in N_p} |p_i - \overline{p}|^2 = \lambda_0 + \lambda_1 + \lambda_3 \qquad (2)$$

Pauly et al.[20] defines the Surface Variation(SV) $\sigma_n(p)$ on the local surface that can quantitatively measure the variation.

$$\sigma_n(p) = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2} \qquad (3)$$

Where $\lambda_0 \le \lambda_1 \le \lambda_2$. If $\sigma_n(p)$=0, then all the vertices are located on a plane. When the neighborhood are the vertices with the isotropically distribution, $\sigma_n(p)$ can attain the maximum 1/3.

Based on (3), SV of every vertices of the 3D ear surface can be computed and are shown in Figure 4 with different colors. It is observed that SV of the key structures of 3D ear such as helix, antihelix, tragus, concha etc. have high values, therefore, selecting the appropriate σ_t, the expected feature vertices p could be detected based on the principle $\sigma_n(p) > \sigma_t$ .In the following, 3-ring neighborhood is used, where $\sigma_t = 0.02$ in experiment 1 and experiment 2, 2-ring neighborhood is used, where $\sigma_t = 0.04$ in experiment 3.

3. 3D ear recognition based on auricle structural feature

The proposed 3DASF include the key physical ear parts and greatly reduces the dimensionality compared to the original data. In this section, the 3D ear recognition based on 3DASF is described, 3DASF is first used to coarsely align the probe gallery ear pairs, then fine alignment is performed with ICP.

3.1 Coarsely align the probe gallery ear pairs based on 3DASF

Let 3DASF point set of the probe ear be M, 3DASF point set of the gallery ear be U, the ICP algorithm is performed between the feature point sets in order to obtain the spatially closest points, which lead to rigid transformation α making error between two points sets minimum. The initial transformation α^0 is set to be the translational transformation of the centroids of two point sets.
The centroids of two point sets are given by

$$C_M = \frac{1}{n}\sum_{i=1}^{n} m_i\ , \quad C_U = \frac{1}{n}\sum_{i=1}^{n} u_i \tag{4}$$

then,

$$\alpha^0 = C_M - C_U \tag{5}$$

The algorithm starts with the initial transformation α^0, and iteratively calculates α^k until converges. In each iteration, the correspondences between two point sets are obtained by looking for the closest points, then the mean square error(MSE) $E(\alpha^*)$ between two point sets is minimized to reach a certain threshold.

The k-th iteration process of ICP algorithm:

(1) Find the closest point set $U^k = \{u_1^k, u_2^k, \ldots, u_n^k\}$ of the point set M^k;

(2) Using SVD to calculate rigid transformation α^k, make $\alpha^k(M^k)$ closest to U^k, i.e. minimizing $E(\alpha^k) = \frac{1}{n}\sum_{i=1}^{n}\left\|\alpha^k(m_i^k) - u_i^k\right\|^2$;

(3) Applying rigid transformation α^k to point set M^k to get point set M^{k+1}, i.e. $M^{k+1} = \alpha^k(M^k)$;

(4) If the difference of MSE $E(\alpha^*)$ between two iterations is below a threshold or exceed a limit number of iteration, i.e. $E(\alpha^{k-1}) - E(\alpha^k) < \tau$, the iteration stop; Otherwise, goto step1 and $k+1$-th iteration continue.

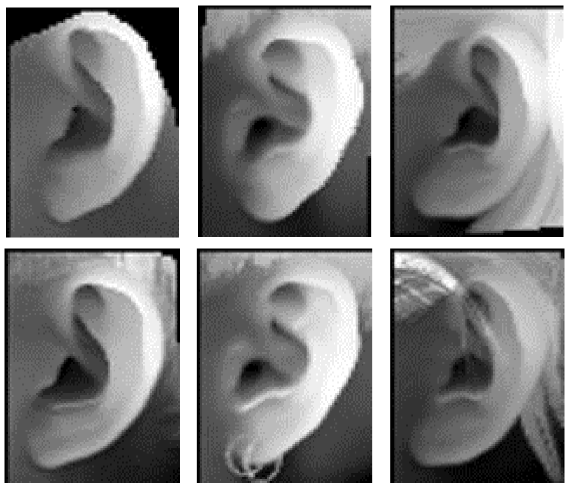
By performing ICP algorithm above, rigid transformation α and MSE $E(\alpha)$ between two closest point sets can be obtained, where α can regarded as the initial transformation of the subsequent ICP algorithm of fine alignment. The attempt to do ear recognition using only $E(\alpha)$ is also tested, which is described in section 4.

3.2 The ICP fine alignment of 3D ear

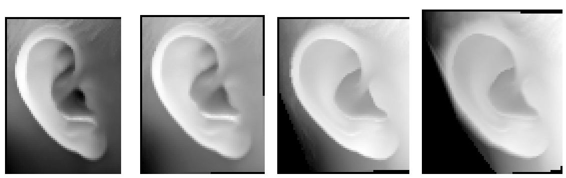
The obtained rigid transformation is used as the initial transformation which is applied to the probe gallery ear pairs, then the ICP algorithm is used to match the probe ear and gallery ear. The gallery ear which has the minimum error with the probe ear will be considered to be the recognition result.
An important step of ICP algorithm is to find the spatially closest point, however, a point in point set X may have one or multi closest point in points set Y. Even if the distance is the shortest, the point may not be the real closest point. Thus, during the process of calculating the rigid transformation, only the point whose distance to its correspondence is less than a threshold τ is selected to calculate the rigid transformation. Generally, τ is set to be the average distance between two point sets plus the double R, R is the resolution of the surface of probe ear.
For the reason that the oiliness of human skin, sensor error and so on, the 3D data from 3D scanner inevitably contains noise, which will lead to "many to one" in the process of finding the closest point. Therefore, the distances between corresponding points are calculated and sorted to select the minor 75% of the distance to update the average distance of two point sets, after several iterations, the effect of noise points will gradually be eliminated.
Finally, the alignment errors are ranked to choose the gallery ear with the smallest error as the recognition result.

4. Experiment

The experimental data comes from UND biometrics datasets collection F and collection G, total of 1,800 3D point cloud data (with the corresponding color images). The data is from 415 individuals, each with 2 or more than 2 (2D color images and the corresponding 3D data), including 237 males, 178 females, there are 70 people with earrings, 40 people ear slightly obscured by hair. A subset of collection G is comprised of 24 people whose images are taken at four different poses, 15° off center, 30° off center, and 45° off center and used to evaluate the impact of pose variation on the performance of recognition algorithm. Figure 5 shows some cases of 3D depth map.
In UND biometrics datasets, collection F and collection G, 2D color image and 3D depth map is registered, so we use Adaboost method [21] to automatically detect and locate ear area in 2D color images, and crop the corresponding area in 3D depth map. In a few cases of automatically ear detection not working, we manually crop ear area. The



(a)Pose (b) Earing (c) Occlusion



(d)Pose variation

Fig. 5 Ear range images

cropped 3D ears are divided into two sets: galley set 1 and probe set 1, each set has 415 point clouds randomly selected from 415 individuals. In the pose variation subset of 24 subjects, the subjects with same pose make up a micro subset. Every image of each micro subset will be matched against all the images from other micro subsets.

Except the adaboost method, all other algorithms are implemented on Intel Xeon (R) CPU 2.4GHz, 12.0G RAM, Maltab 7.12.0 (R2011a).

In order to validate the proposed 3DASF for ear, we first try to only use 3DASF to carry out 3D ear recognition on galley set 1 and probe set 1. The ICP algorithm described in section 3.1 is used to align the 3DASF points of probe ear and gallery ear. The candidate ear which has the minimum MSE with probe ear will be considered as the recognition result. Cumulative Match Characteristics (CMC) is used to evaluate the performance. The results is shown in Table 1. The limit number of ICP algorithm iteration is set to be 30 times, iterative threshold $\tau = 1 \times e^{-4} mm$.

Table 1: CMC of only using 3DASF

Rank-1	Rank-2	Rank-3	Rank-4	Rank-5
77.59%	79.52%	82.17%	82.41%	82.89%

The rank-1 recognition rate is only 77.59% when 3DASF is used alone. The reason can be summed up that the 3DASF usually contain noise points of the background area, when background of probe ear and gallery ear change greatly, the error will be sharply increased that degrade the recognition rate.

Table 2 shows the CMC of coarsely aligning the probe gallery ear pairs, and then ICP fine alignment is performed on galley set 1 and probe set 1. Since the extracted 3DASF of ear contains important physiological region, a good initial rigid transformation can be obtained after coarsely aligning the probe gallery ear pairs. Then, ICP algorithm is used for fine alignment, that can get a better recognition performance compared with directly using ICP fine alignment. The Rank-1 recognition rate is 97.59%. The limit number of ICP algorithm iteration is set to 30 times, iterative threshold $\tau = 1 \times e^{-4} mm$.

Table 2: CMC of "coarsely aligning the probe gallery ear pairs with 3DASF + ICP fine alignment"

Rank-1	Rank-2	Rank-3	Rank-4	Rank-5
97.59%	98.55%	99.04%	99.04%	99.04%

In order to solve the pose problem which can not be overcame with 2D recognition approach, the proposed 3D ear recognition method based on 3DASF is evaluated on the pose variation subset for robustness on the occasion of pose variation. The four micro subsets are considered as the gallery set and the probe set in turns, and cross matched against each other. The results are shown in Table 3. It can be observed that the results obtained by the proposed method are better than Hui[11]'s, Yan[22]'s and Zhou[15]'s at the mean value of rank-1 rate. The relative invariability of the 3DASF is shown and the proposed method is with robustness to pose variation to some extent.

The comparison of our method and other 3D ear recognition method is shown in Table 4. The proposed method needs less matching time and achieves higher preferred recognition rate than other method which are based on ICP algorithm. Compared with Zhou's method, although its match time and the recognition rate are better, but the extraction process of SPHIS local feature is relatively complex, involving feature point selection, surface curvature calculation and so on, thus its computation time can not be ignored. Zhou did not mentioned the time cost of SPHIS local feature extraction. With the same principle with Zhou' work, Islam' L3DF can be used as a reference to SPHIS. L3DF ear feature extraction time for a single person is 22.2 seconds. By comparison with that, the proposed 3DASF extraction time is 1.11 seconds, which has an obvious advantage.

Table 3 Comparison among this paper's method and others on the subset with pose variation of Collection G

Gallery / Probe	90°	75°	60°	45°	Average
90°		**100%** [100%, 100%, 100%]	**95.8%** [87.5%, 87.5%, 87.5%]	**87.5%** [83.3%, 70.8%, 70.8%]	**94.4%** [90.3%, 86.1%, 86.1%]
75°	**100%** [100%, 100%, 100%]		**100%** [100%, 100%, 100%]	**91.7%** [91.7%, 87.5%, 83.3%]	**97.2%** [97.2%, 95.8%, 94.4%]
60°	**100%** [91.7%, 87.5%, 91.7%]	**100%** [100%, 100%, 100%]		91.7% [91.7%, **95.8%**, 87.5%]	**97.2%** [94.4%, 94.4%, 93.1%]
45°	83.3% [**87.5%**, 79.2%, 58.3%]	**91.7%** [87.5%, 87.5%, 83.3%]	95.8% [87.5%, **100%**, 83.3%]		**90.3%** [87.5%, 88.9%, 75%]
Average	**94.4%** [93.1%, 88.9%, 83.3%]	**97%** [95.8%, 95.8%, 94.4%]	**97.2%** [91.7%, 95.8%, 90.3%]	**90.3%** [88.9%, 84.7%, 80.5%]	**94.8%** [92.4%, 91.3%, 87.1%]

Hui[11]'s, Yan[22]'s and Zhou[15]'s results are given in the brackets. The better recognition results are bolded.

Table 4: Comparison with other 3D ear recognition methods

Authors	Preprocess	Methods	Time of feature extraction(one person)	Matching time	Rank-1
Yan[22]	Skin detection + Ear hole location + Active Contour Model	ICP(415 person)	N/A	5~8 s	97.6%
Chen[11]	Skin detection + Edge extraction + Reference Ear Shape Model	LSP + ICP(302 person)	3.7 s		96.36%
Zhou[15]	HIS ear location	SPHIS+ Voxelization(415 person)	Not reported	0.02 s	98.3%
Islam[14]	Adaboost ear detection	L3DF+ICP(415 person)	22.2 s	2.28 s	93.5%
This paper	Adaboost ear detection	3DASF+ICP(415 person)	1.11 s	0.22 s	97.59%

5. Conclusion

This paper propose a new 3D ear representation for recognition--3DASF and the corresponding 3D ear recognition method. The 3DASF extraction is based on the estimating the variation of mesh surface, which contains the important information about ridge and channel of ear. 3DASF is used for coarsely alignment and further ICP algorithm is used for fine alignment. The better recognition performance than the existing recognition method based on ICP algorithm is achieved. The experiment on UND biometrics datasets collection F and collection G from 415 individuals shows that the rank-1 recognition rate is 97.59%, and feature extraction and matching time is respectively 1.11 seconds and 0.22 seconds. The experiment on the pose variation subset of collection G shows the robustness of the proposed method to pose variation.

The low recognition rate when only using 3DASF is discovered in the experiment. Therefore, in the future work, how to overcome the effect of background noise, extract more accurate and efficient 3D structure feature to form a better recognition solution with flexibility is a open problem.

Acknowledgments

The authors are grateful to computer vision research laboratory at University of Notre Dame for providing UND biometrics datasets, collection F and collection G.
This work is supported the Natural Science Foundation of China under grant no. 60973064, the Doctoral Fund of Ministry of Education of China under grant no. 20100006110014, and the Beijing Municipal Natural Science Foundation under grant no. 4102039

References

[1] M. K. J. R.Rajalakshmi, "A review on classifiers used in face recognition methods under pose and illumination variation", International Journal of Computer Science Issues, vol.9, no.6 2012, pp. 474~485.
[2] C. Kyong, K. W. Bowyer, S. Sarkar and B. Victor, "Comparison and combination of ear and face images in appearance-based biometrics", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol.25, no.9 2003, pp. 1160-1165.
[3] P. J. Besl and N. D. McKay, "A method for registration of 3-d shapes", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol.14, no.2 1992, pp. 239-256.
[4] Y. Ping and K. W. Bowyer, "Ear biometrics using 2d and 3d images", in Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on, 2005, pp. 121-121.
[5] P. Yan and K. W. Bowyer, "A fast algorithm for icp-based 3d shape biometrics", Proc. Proceedings of the Fourth IEEE Workshop on Automatic Identification Advanced Technologies2005, pp. 213-218.
[6] P. Yan and K. W. Bowyer, "Empirical evaluation of advanced ear biometrics", in Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on, 2005, pp. 41-41.
[7] P. Yan, K. W. Bowyer and K. J. Chang, "Icp-based approaches for 3d ear recognition", 2005, pp. 282-291.
[8] H. Chen, B. Bhanu and R. Wang, "Performance evaluation and prediction for 3d ear recognition", in Audio- and video-based biometric person authentication, eds. T. Kanade, A. Jain and N. Ratha, pp. 748-757, Springer Berlin Heidelberg, 2005.
[9] P. Yan and K. W. Bowyer, "An automatic 3d ear recognition system", in Proc. Third Int'l Symp. 3D Data Processing, Visualization, and Transmission, 2006, pp. 213-218.
[10] H. Chen and B. Bhanu, "3d free-form object recognition in range images using local surface patches", Pattern Recogn Lett, vol.28, no.10 2007, pp. 1252-1262.
[11] C. Hui and B. Bhanu, "Human ear recognition in 3d", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol.29, no.4 2007, pp. 718-737.
[12] C. Hui and B. Bhanu, "Efficient recognition of highly similar 3d objects in range images", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol.31, no.1 2009, pp. 172-179.
[13] K. Wang, Z. Mu and Z. He, "An efficient approach of 3d ear recognition", in 10th World Congress on Intelligent Control and Automation, WCICA 2012, July 6, 2012 - July 8, 2012, 2012, pp. 4784-4790, Institute of Electrical and Electronics Engineers Inc., Beijing, China.
[14] S. Islam, R. Davies, M. Bennamoun and A. Mian, "Efficient detection and recognition of 3d ears", International Journal of Computer Vision, vol.95, no.1 2011, pp. 52-73.
[15] J. Zhou, S. Cadavid and M. Abdel-Mottaleb, "An efficient 3-d ear recognition system employing local and holistic features", IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, vol.7, no.3 2012.
[16] http://www.nd.edu/~cvrl/.
[17] http://www.mathworks.de/matlabcentral/fileexchange/8998.
[18] M. Pauly, M. Gross and L. P. Kobbelt, "Efficient simplification of point-sampled surfaces", in Visualization, 2002. VIS 2002. IEEE, 2002, pp. 163-170, IEEE.
[19] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald and W. Stuetzle, "Surface reconstruction from unorganized points", SIGGRAPH Comput. Graph., vol.26, no.2 1992, pp. 71-78.
[20] M. Pauly, M. Gross and L. P. Kobbelt, "Efficient simplification of point-sampled surfaces", Proc. Proceedings of the conference on Visualization '02, Boston, Massachusetts2002, pp. 163-170.
[21] L. Wen-Jing and M. Zhi-Chun, "Ear detection based on improved adaboost algorithm", in Pattern Recognition, 2008. CCPR '08. Chinese Conference on, 2008, pp. 320-325.
[22] P. Yan and K. W. Bowyer, "Biometric recognition using 3d ear shape", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.29, no.8, Aug 2007, pp. 1297-1308.

Kai Wang received the B.E. degree from the School of Electrical Engineering at Henan Polytechnic University in 2005. He is currently pursuing Ph.D. in Control Science and Control Engineering from the School of Automation and Electrical Engineering, University of Science and Technology Beijing. His research interests include computer vision, pattern recognition, biometrics and machine learning.

Zhi-Chun Mu received the B.E. and M.E. degrees from the Department of Automation at University of Science and Technology Beijing in 1978 and 1983, respectively. He is currently a professor in the School of Automation and Electrical Engineering, University of Science and Technology Beijing. He was a visiting scholar at Davy International UK and Sheffield Polytechnic, England, from 1989 to 1991, and a visiting Professor at University of Brighton, England from 1996 to 1999. As a Guest Professor, he visited Laboratoire d'Electronique, d'Informatique et d'Image, CNRS University of Burgundy France in 2007. He was the Chair of Organizing Committee of IEEE International Conference on Wavelet Analysis and Pattern Recognition 2007. He has served as
a reviewer and member of Evaluation and Assessment Commission, Division of Information Science, National Natural Science Foundation of China since 2002.He is now Chair of IEEE SMC Beijing (capital region) Chapter. His main research interests include Pattern Recognition and Biometrics, Artificial Intelligence and its applications, Data Mining as well as Process Control and Modeling. He has published more than 180 refereed journal and conference papers in these areas. He is the corresponding author of this paper.

Frequency-Domain Modeling and Simulation of Coupled Lossy Multiconductor Transmission Lines

Youssef Mejdoub[1], Hicham Rouijaa[1,2], Mohamed Saih[1], and Abdelilah Ghammaz[1]

[1] Laboratory of Electrical Systems and Telecommunications, Department of Physics, Faculty of Sciences and Technology, Cadi Ayyad university, P.O. Box 549/40000, Marrakesh, Morocco

[2] Department of Material Sciences, Poly-disciplinary Faculty, Cadi Ayyad university, P.O. Box 4162/46000, Safi, Morocco

Abstract

This paper presents a study of modeling and simulation of the lossless and lossy multiconductor transmission lines in frequency domain. This study is based on a characteristics method, which permits modeling the line as a quadripole whose advantage is not to presuppose applied charge conditions in its extreme. This permits it to be introduced easily in the circuit simulators such as Spice, Esacap and Saber. The results presented here cover two types of multiconductor transmission lines, the "Ribbon Cable" and "2Wire Xtalck.

Keywords: *Characteristics method, Frequency domain, Multiconductor transmission line, Losses.*

1. Introduction

The problems related to the frequency mounting and the effects of interconnections in equipments and industrial applications are various (distortion, attenuation, crosstalk, etc...). Moreover the losses can play a very important role in the degradation and attenuation of signals moving through the line. Considering the framework of the inter-equipment electromagnetic compatibility (EMC), skin and proximity effects become more complicated in high frequency, which necessitates, therefore, a physical model adapted to the transmission lines.

The cascade model [1][2] allows modeling the transmission line in the form of a RLCG circuit. It can also be applied in case of MTL line, which becomes quite complicated when using more than two conductors. And it results not only in oscillations in time domain (phenomenon of Gibbs), but it needs an important calculating time, which makes it inefficient.

The finite differences method in time domain (FDTD) [2][3][4] is an analytic method. It consists of dividing the time and the space where the solution is searched for in a network of point spaced regularly to constitute a mesh.

The major difficulty then, is to have a distributed MTL line model valid in both time and frequency domains, with and without losses. The model presented in this paper for modeling and simulation of the lossless and lossy multiconductor transmission lines in frequency domain, it's valid also in the time domain [5].

This paper will expose a behavior study of the lossless and the lossy multiconductor transmission line –MTL- in frequency domain, based on the modelling of the MTL lines numerical method with the help of Branin model, that can analyze the MTL line by representing it in the form of a quadripole, this is the characteristics method [2][6][7][8], This method would permit the analysis of an MTL line without and with losses and present the advantage to avoid presupposing conditions of applied charges to its extremes. This permits it to be introduced easily in circuit simulators as Spice, Esacap and Saber.

Diverse examples of applications are presented to validate this model and to show their interests and to understand the behavior of a line MTL in the frequency domain.

2. MTL Lines Modeling

2.1 Lossless Lines

Branin [6] was the first to propose a numerical model of a transmission line which permitted giving a simple schema equivalent to the ideal line. This schema includes two dipoles. In the input dipole (z=0,t) the tension is determined from the reflected tension in extremes of the line (z=l) at the precedent time t-Tr. The same interpretation is applied to the output dipole (Tr is the delay of the line).

Following [2][6], the equivalent schema of the lossless line (R=G=0) is depicted in the fig.1 :

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 1, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

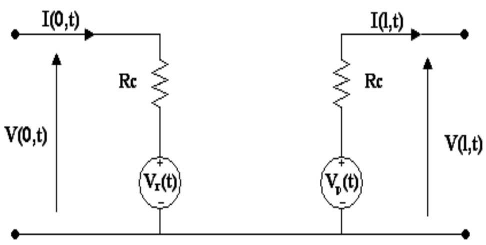


Fig.1 : Quadripole representation of the ideal line.

Indeed, it is easy to show that electrical parameters in extremities are linked by means of the following relations:

$$\begin{aligned} V(0,t) &= R_C I(0,t) + V(l,t-T_r) + R_C I(l,t-T_r) \quad (a) \\ V(l,t) &= -R_C I(l,t) + V(0,t-T_r) + R_C I(0,t-T_r) \quad (b) \end{aligned} \tag{1}$$

The equations (1.a) and (1.b) do not depend only on secondary parameters (the characteristic impedance and the delay of the line). Hence, the equivalent schema of the figure 1 is established by putting:

$$\begin{aligned} V_r(l,t) &= V(l,t-T_r) + R_C I(l,t-T_r) \quad (a) \\ V_i(0,t) &= V(0,t-T_r) + R_C I(0,t-T_r) \quad (b) \end{aligned} \tag{2}$$

$V_r(t)$ *and* $V_i(t)$ are then calculated using parameters observable at a known time $t - T_r$.

The features method can perhaps be spread to the multiconductor lines case [2][7][9][10]. This requires, on the other hand, uncoupled propagation modes on each line. This separation of modes is implemented using the modal method.

Keeping in mind electromagnetic couplings between the lines, the impedances matrix are not diagonal. The interest of the modal method is then to uncouple the equations to be able to diagonalise the matrix.

To model the transmission lines, it is necessary to solve the telegraphists equations. The time domain equations representation is:

$$\begin{cases} \dfrac{\delta V(z,t)}{\delta z} + RI(z,t) + L\dfrac{\delta I(z,t)}{\delta t} = 0 \\ \dfrac{\delta I(z,t)}{\delta z} + GV(z,t) + C\dfrac{\delta V(z,t)}{\delta t} = 0 \end{cases} \tag{3}$$

[V] and [I] represent, respectively, the tensions and currents matrix.

[R], [L], [C] and [G] represent, respectively, the resistances, inductances, conductance's and capacities matrix. These include implicitly of all informations concerning the transverse section, which permits to characterize a muticonductor structure.

The coefficients of these different matrix are obtained either by engineering practical techniques [11], or by numerical methods [9][10][12]. The modelling of a lossless and lossy MTL lines in frequency domain by the Branin method will be exposed respectively in paragraphs A and B.

As we deal with the lossless lines case (G=R =0), the MTL lines equations in the time domain (3) become:

$$\begin{cases} \dfrac{\delta^2 V(z,t)}{\delta z^2} - LC\dfrac{\delta^2 V(z,t)}{\delta t^2} = 0 \\ \dfrac{\delta^2 I(z,t)}{\delta z^2} - CL\dfrac{\delta^2 I(z,t)}{\delta t^2} = 0 \end{cases} \tag{4}$$

The modal method introduces fictitious parameters V_m *and* I_m, by means of the following linear transformations:

$$\begin{cases} V(z,t) = T_V . V_m(z,t) \\ I(z,t) = T_i . I_m(z,t) \end{cases} \tag{5}$$

The modal matrix T_V *and* T_I, are to be determined to ensure the uncoupling of propagation modes of the MTL line [2][7][9][10]. If we apply the transformation to the system of equations (4), we then obtain:

$$\begin{cases} \dfrac{\delta^2 V_m(z,t)}{\delta z^2} - T_V^{-1} LCT_v \dfrac{\delta^2 V_m(z,t)}{\delta t^2} = 0 \\ \dfrac{\delta^2 I_m(z,t)}{\delta z^2} - T_i^{-1} CLT_i \dfrac{\delta^2 I_m(z,t)}{\delta t^2} = 0 \end{cases} \tag{6}$$

We can choose both matrix T_V *and* T_I in such way that the system (6) could be uncoupled. In our case of a lossless line placed in a homogeneous and isotropic medium, we can always find the transformation matrix which diagonalise simultaneously the matrix L and C.

The two transformation matrix T_V *and* T_I are bound by the following relation [2]:

$$T_i^t = T_V^{-1} \tag{7}$$

The system of equation of a lossless line spells in the modal base :

$$\begin{cases} \dfrac{\delta V_m(z,t)}{\delta z} = -L_m \dfrac{\delta I_m(z,t)}{\delta t^2} \\ \dfrac{\delta I_m(z,t)}{\delta z} = -C_m \dfrac{\delta V_m(z,t)}{\delta t} \end{cases} \tag{8}$$

With L_m *and* C_m are diagonal matrix of dimension NxN

$$\begin{cases} L_m = T_V^{-1} . L . T_i \\ C_m = T_i^{-1} . C . T_V \end{cases} \tag{9}$$

The system of equations (8) represents an uncoupled MTL line, which has a characteristic impedance R_{Cm} and a delay T_{rm}:

$$R_{Cm_{ii}} = \sqrt{\frac{L_{m_{ii}}}{C_{m_{ii}}}}, \quad T_{rm_i} = \frac{l}{\sqrt{L_{m_{ii}} C_{m_{ii}}}} \qquad (10)$$

2.2 Lossy Lines

The losses can play a very important role in the degradation and attenuation of signals transported by line. The losses are caused both by a non-null conductivity and of the loss of the polarization medium or imperfect conductors. The losses presented by imperfect conductors are habitually more significant than those owing to the transmission line. For this reason, we often suppose that the surrounding medium is lossless (G = 0) in the MTL line equations. The resistance due to imperfect conductors is represented in the resistance matrix by a length unity [R].
The system of equations (3) becomes:

$$\begin{cases} \frac{\delta V(z,t)}{\delta z} + RI(z,t) + L\frac{\delta I(z,t)}{\delta t} = 0 \\ \frac{\delta I(z,t)}{\delta z} + C\frac{\delta V(z,t)}{\delta t} = 0 \end{cases} \qquad (11)$$

The equations (11) expressed in the frequency domain by Laplace operator p gives us:

$$\begin{cases} -\frac{\delta V}{\delta z} = [R][I] + p[L][I] = [Z][I] \\ -\frac{\delta I}{\delta z} = p[C][V] = [Y][V] \end{cases} \qquad (12)$$

In general, a line can be modeled by its serial impedance $Z = R + jwL$, and parallel admittance $Y = jwC$, (The dielectrics losses are supposed to be negligible G=0).
For the working frequencies well superior to a characteristic frequency of the line equal $R/2\pi L$ (low losses hypothesis) [2][7], and using a first development order, we obtain:

$$Z_C = \sqrt{\frac{R + jwL}{iwC}} \approx R_C + \frac{RR_C}{2jwL} \quad where\, R_C = \sqrt{\frac{L}{C}} \qquad (13)$$

The characteristic impedance in that case is equivalent to characteristic impedance RC mounted serially with a capacity $C_{pf} = \frac{2L}{R.R_C}$, when the frequency increases ZC becomes equal to RC.

With the same approximation, the constant of propagation becomes: $\gamma = \alpha + j\beta = \frac{R}{2R_C} + jw\sqrt{LC}$

To take into account the attenuation of the wave the term $e^{-\alpha l}$ is introduced, it is enough, then, to modify Branin's generators of the quadripole representation of the ideal line (cf Fig 2). They become [2][7] :

$$\begin{cases} V_r'(t) = V_r(t).e^{-\alpha l} \\ V_i'(t) = V_i(t).e^{-\alpha l} \end{cases} \qquad (4)$$

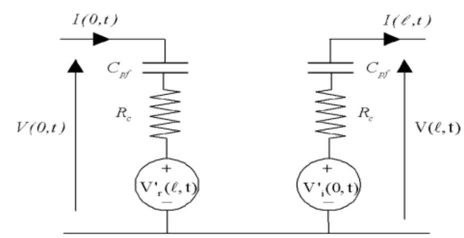


Fig.2 : Quadripole representation of the low loss line.

For a lossy multiconductor line, we do the same as in the case of a lossless MTL line. Therefore, in the case of coupled low loss lines, the characteristic impedance and the delay are given by:

$$Z_{Cm} \approx R_{Cm} + \frac{RR_{Cm}}{2jwLm} \quad , R_{Cm} = \sqrt{\frac{L_m}{C_m}} \qquad (15)$$

Thais supposes that the resistance's losses referred to the R matrix whose non-diagonal terms were null. Also, it is enough to bring the following modifications to the generators of the ideal lines, given by relations (2.a) (2.b), which become:

$$\begin{aligned} V_{rm}(l,t) &= T_V'\{V_m(l,t-T_{rm}) - Z_{Cm}I_m(l,t-T_r)\} \quad (a) \\ V_{im}(0,t)' &= T_V'\{V_m(0,t-T_{rm}) + Z_{Cm}I_m(0,t-T_r)\} \quad (b) \end{aligned} \qquad (16)$$

With $T_V' = T_V e^{-\alpha l}$ is the attenuated modal matrix.
The Electric parameters to the extremes of the MTL line are linked by the following relations:

$$\begin{aligned} V_m(l,t) &= -Z_{Cm}I_m(l,t) + V_{im}'(0,t) \quad (a) \\ V_m(0,t) &= Z_{Cm}I_m(0,t) + V_{rm}'(l,t) \quad (b) \end{aligned} \qquad (17)$$

3. Simulation & Results

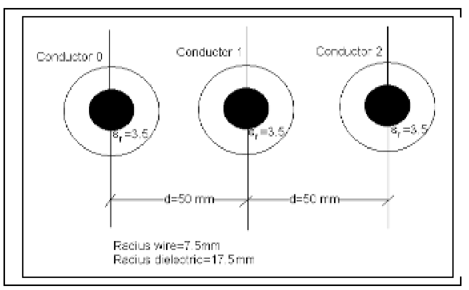
To verify the validity of this model in frequency domain, we simulate two types of multiconductor transmission lines, the "Ribbon Cable" and "2Wire Xtalck in ESACAP circuit simulator [13]. In this paragraph, we make a comparison between simulation results in the literature for example 1[2] and example 2 [10] and our simulation results using the ESACAP circuit simulator in frequency domain in the lossless and lossy. This model is valid also in the time domain [5].

3.1 Example 1: Ribbon Cable

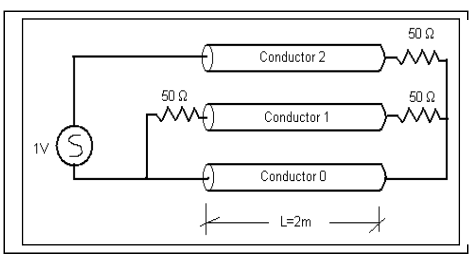
We consider a transmission line of three conductors with 2m of length and ligneous parameters R,L and C, fed by a tensional generator, the in and the out charges are equal to 50 Ohms, as indicated on the figure 3.

$$L=\begin{bmatrix}750.37 & 508.69\\ 508.69 & 750.37\end{bmatrix}(nH/m) \qquad R=\begin{bmatrix}0.19 & 0\\ 0 & 0.19\end{bmatrix}(\Omega/m)$$

$$C=\begin{bmatrix}37.21 & -18.60\\ -18.60 & 37.21\end{bmatrix}(pF/m)$$

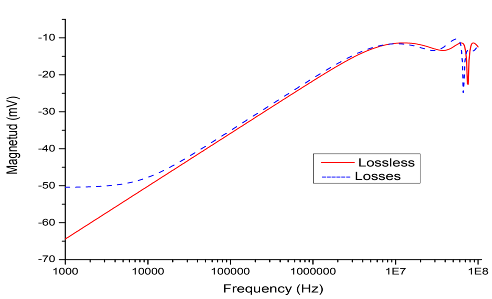


a) dimensions.

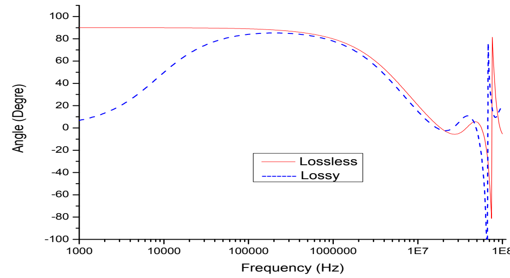


b) Electric representation.

Fig.3 : Ribbon Cables transmission line to 3 conductors.

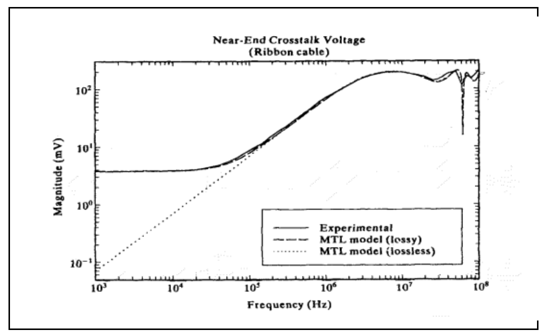


a) Magnitude.

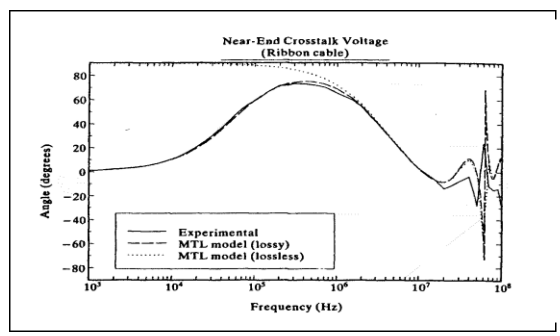


b) Phase.

Fig.4 : The near-End crosstalk voltage simulated with Esacap 2000.



a) Magnitude.



b) Phase.

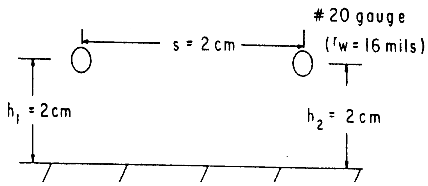
Fig.5 : Validation result of the near-End crosstalk voltage [2].

Simulation results: The experimental results from Ref [2] are compared to the predictions of our method, over frequency range of 1 kHz to 100 MHz. As from figure (4.a), for F<100 kHz the losses in the line conductors are predominant. For F>100KHz, the inductive effect becomes predominant over resistive effect.
In the Figure 4, the amplitude and the phase of near-end crosstalk, the predicted voltage induced at the source side of conductor n.1, are displayed. The corresponding values from Ref. [2] are displayed for comparison on figure 5.
As example, the low-frequency value in Figure (4.a) is ~ -48dBV, corresponding to ~ 4 mV in Figure (5.a).

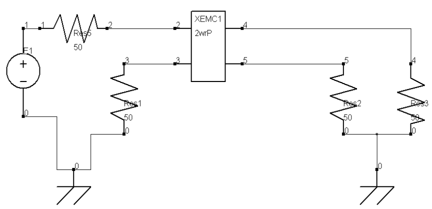
Physical interpretation: From (4.a) at low frequency (R-DC >wL), the resistive coupling is predominant, while at intermediate frequency, the classic inductive X-talk becomes predominant (slope +20dB/dec).
At high frequency, the voltage is limited by the line's characteristic impedance. The sharp resonance at F~65MHz is the λ/2 resonance of 2m cable, shifted down by increased electrical length due to extra capacitance of dielectric gains. The λ/2 resonance is excited because of low R terminations, compared to common mode impedance Zcm~ 130Ω.

3.2 Example 2: '2Wire Xtalk'

We consider a '2wire-xtalk' transmission line with length 4.674m and of dielectric constant ε =1, fed by a tensional generator, the in and the out charges are equal to 50 Ohms, as indicated on the figure 6.



a) Geometric configuration



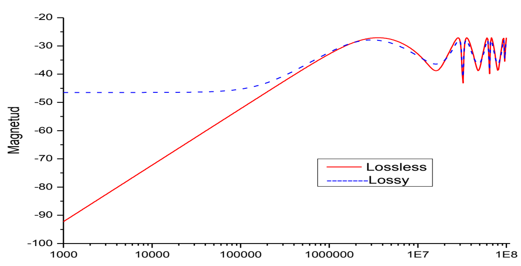
b) Electric model of the line

Fig.6 : configuration of the line '2wire-xtalk'.

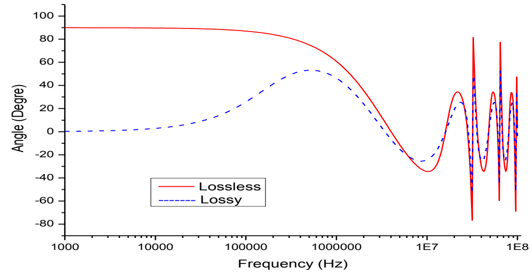
Figure 6 presents the geometric configuration of the line whose parameters are:

$$L = \begin{bmatrix} 918 & 161 \\ 161 & 918 \end{bmatrix} (nH/m) \quad R = \begin{bmatrix} 0.78 & 0 \\ 0 & 0.78 \end{bmatrix} (\Omega/m)$$

$$C = \begin{bmatrix} 12.50 & -2.19 \\ -2.19 & 12.50 \end{bmatrix} (pF/m)$$

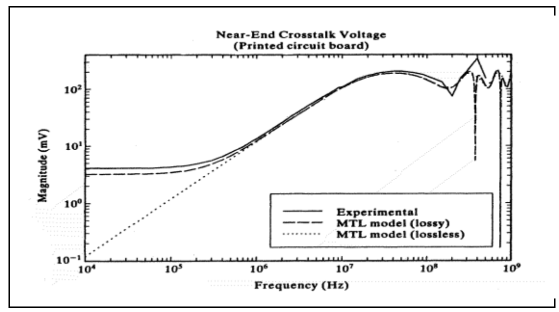


a) Magnitude

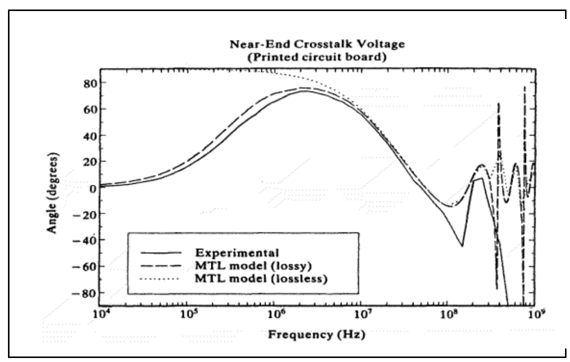


b) Phase

Fig.7 : The near-End crosstalk voltage simulated with Esacap 2000



a) Magnitude



b)

b) Phase

Fig.8 : Validation result for the near-End crosstalk voltage [2].

Simulation results: The predictions of our method are compared to experimental results (Ref.[2]), for frequency range F=10 kHz -1GHz. At F<200KHz, the conductor losses become predominant.
In the following figures, the amplitude and phase of near-end crosstalk voltage on conductor n. 1 are shown, for comparison between EMCAP prediction results and reference canonical results (Ref. [2]).
For example, the low-frequency value in Fig.(7.a) is -49.7dBV, equivalent to 3.3 mV in the figure (8.a).

Physical interpretation: The same considerations of previous example in the Section 3.2 are applicable. The frequency of the sharp resonance is now shifted to F~400MHz, compared to previous one F~ 65MHz, with scaling almost similar to length ratio (2m against 0.25 m) . The increased capacitance due to PCB substrate, compared to wire dielectric gains, explains the difference between the frequency and length ratios.

4. Conclusion

In this work, we show that the interest of the characteristics method modeling the MTL lines in frequency domain. This method can be introduced easily in circuit simulators as Spice, Esacap and Saber, but it is only valid in low loss MTL line in the time and the frequency domains. The future work will be devoted to the development model of both real losses using a model based on the Pade approximation, and the effect of disturbing EM wave disturbed on a MTL line.

References

[1] C.W. HO, "Theory and computer aided analysis of lossless transmission lines", IBM. J. Res. Dev, Vol. 16, No. 3, 1973, pp. 249-255.

[2] C. R. Paul, Analysis of Multiconductor Transmission Lines, Wiley series in Microwave and Optical Engineering, Kai Chang, Series Editor, 1994.

[3] A. ASSADI-HAGHI, "Contribution au développement de méthodes d'optimisation structurelle pour la conception assistée par ordinateur de composants et de circuits hyperfréquences", Ph.D. Communications Optiques et Microondes, Université de Limoges Ecole Doctorale Sciences et Techniques Faculté des sciences et Techniques XLIM- Département Minacom, 2007.

[4] K. Afrooz, A. Abdipour, A. Tavakoli and M. Movahhedi, "Time-domain analysis of lossy active transmission lines using FDTD method", Int. J.Commun. (AEU), Vol. 63, No. 3, 2009, pp. 168-178.

[5] Y. Mejdoub, H. Rouijaa and A. Ghammaz, "Transient analysis of lossy multiconductor transmission lines model based by the characteristics method", International Journal of Engineering and Technology IJET, Vol. 3, No. 1, 2011, pp. 47-53.

[6] F. H. Branin, "Transient Analysis of Lossless Transmission Lines", Proc. IEEE, Vol. 55, No. 11, 1967, pp.2012-2013.

[7] H. ROUIJAA, "Modélisation des Lignes de Transmission Multiconducteurs par La méthode des Approximantes de Pade : Approche circuit", Ph.D. Génie Electrique, Université de Droit d'Economie et des Sciences d'Aix-Marseille (Aix-Marseille III), Mai 2004.

[8] A. Dounavis, A. Pothiwala and A. Amir Beygi, "Passive macromodels of lossy multiconducteur transmission lines based on the method of characteristics", IEEE Trans. Microwave Theory, Vol.32, No. 1, 2009, pp.184-198.

[9] M. Kane, Ph. Seltner and Ph Auriol, "Détermination des paramètres linéiques des câbles multifilaires", in 6ème Colloque International CEM-92, 1992,pp. 323-328.

[10] M. KANE, "Modèles analytiques originaux pour la détermination des paramètres linéiques des lignes et câbles multifilaires parcourus par des signaux large bande", Ph.D. Génie Electrique, école Doctorale de Lyon des Sciences pour L'ingénieur : Electronique ; Electrotechnique ; Automatique, 1994.

[11] S. BAZZOLI, "Caractérisation et Simulation de la Susceptibilité des Circuits Intégrés face aux Risques d'Inductions engendrées par des Micro-ondes de Forte Puissance", Ph.D Electronique, Université des Sciences et Technologies de Lille (Lille-1), octobre 2005.

[12] C. R. PAUL, "Introduction to Electromagnetic compatibility", wiley series in Microwave and Optical Engineering, Kai Chang, Series Editor. 1992.

[13] L. Inzoli, H. Roujaa, G. Akoun and J. Robert "Empap2000-Esacap, un outil de simulation intégrant le couplage champ-circuit", in 11ème Colloque international & Exposition sur la CEM , Grenoble, 2002.

Youssef MEJDOUB was born in Morocco, in 1980. He is received the DESA (6 years study after the baccalaureate, equivalent to Master) in Telecommunications & Network from Cadi Ayyad University, Marrakesh Morocco, in 2007. He is a PhD student in Electrical Systems and Telecommunications Laboratory ***LSET*** at Cadi Ayyad University – Marrakech - Morocco. His research interest includes the electromagnetic compatibility, multiconductor transmission lines and telecommunications.

Mohamed SAIH obtained his diploma in Electrical Engineering from Cadi Ayyad University – Morocco in 2011. He is a researcher member of the Electrical Systems and Telecommunications Laboratory ***LSET*** and preparing his doctorate thesis. His research interest includes the electromagnetic compatibility and multiconductor transmission lines.

Hicham ROUIJAA is a Professor of physics, attached to Cadi Ayyad University, Marrakesh Morocco. He is obtained his PhD Thesis on \" Modeling of Multiconductor Transmission Lines using Pade approximant method: Circuit model \", in 2004, from Aix-Marseille University - France. He is associate member of Electrical Systems and Telecommunications Laboratory ***LSET*** at the Cadi Ayyad University. His current research interests concern electromagnetic compatibility and multiconductor transmission lines.

Abdelilah GHAMMAZ received the Doctor of Electronic degree from the National polytechnic Institut (ENSEEIHT) of Toulouse, France, in 1993. In 1994 he went back to Cadi Ayyad University of Marrakech – Morroco. Since 2003, he has been a Professor at the Faculty of Sciences and technology, Marrakech, Morroco. He is a member of ELectrical Systems and Telecommunications Laboratory ***LSET*** at the Cadi Ayyad University. His research interests in the field of electromagnetic compatibility, multiconductor transmission lines, telecommunications and antennas.

Building web crawler based on bee swarm intelligent algorithm

mohammed Ibrahim Shujaa[1] ahmed bahaa ulddin [2],

[1] department of informatics, foundation of technical education
Baghdad, State ZIP/Zone, Iraq
[2] department of informatics, foundation of technical education
Baghdad, State ZIP/Zone, Iraq

Abstract
Search engines are using web spiders to crawl the web in order to collect copies of the web sites for their databases, these spiders usually use the technique of breadth first search which is non-guided (blind) depends on visiting all links of any web site and one by one. This paper proposed a new algorithm for crawling web depending on swarm intelligence techniques, the adopted algorithm is bee swarm algorithm which takes the behavior of the bee for its work, the result in terms of speed and accuracy which means the relevancy of the collected sites.

Keywords: *web crawling, crawlers, bee swarm algorithm, swarm intelligence.*

1. Introduction

One of the most essential jobs of any search engine is gathering web pages also called crawling; web crawler is a software or program that uses the graphical structure of the web to move from page to page and add them to a local database, web crawlers used to create a copy of all visited pages for later processing by search engine], another definition for the web crawler (also known as robot) is a system for downloading of the web pages, web crawlers used for variety of purposes[1].
Most permanently considered one of the main component Of the web search engines that assemble a corpus of web pages, index them and allow users to issue queries against the index and find the web pages that match the queries, a related use is web archiving where large sets of web pages are periodically collected and archived for posterity and web data mining where web pages are analyzed for statistical purposes [2].the simplest form of web crawler starts from a seed page and then uses the external links with it to attend other pages, the process repeated with new pages offering more external links To follow until a sufficient number page-s are collected or some high level objective is reached, the working of crawler can be shown in the following algorithm:

1-start
2-initialize (frontier) query by adding seed URL to it.
3-IF frontier empty THEN go to 7
Else
Pick up a URL from frontier
4-fetch the page corresponding to the URL
5-parse the page to find new URL's
6- Add new unvisited URL's to the queue
7-end

Therefore, any search engine's web crawler should have the following objectives:-
1-it should explore and download web documents from the World Wide Web as much as possible.
2-it should bring a high quality documents so that the users gets the required relevant information within acceptable time.
3-the documents must be displayed in order to their relevance.
4-the web documents is very dynamic search engine should update it's repository.
To satisfy the first objective, search engine depend on multiple crawlers, and the ranking algorithm satisfy the second and third objective, the fourth objective can be satisfied by frequent crawling [1].

2- Problem state

in order to get an accurate functionality of the search engine, it need an efficient web crawler providing an accurate database for web sites, the problem of the web crawling is the crawling of not relevant or not useful pages and slowness of most crawlers, so this paper propose a new web crawling algorithm depends on the bee swarm intelligence algorithm in order to speed up crawling and use the collective intelligence of the bee to crawl Most relevant (useful) web pages with good speed.

3-swarm intelligence

Swarm intelligence is a section of artificial intelligence based on study of actions in various decentralized system, this decentralized system (multi agent system) are composed of the physical individuals (robot) or virtual ones that communicate among themselves, cooperate,

exchange information and knowledge and perform some tasks in their environments [4,5].
Another definition from Dorgio and Birattori swarm intelligence is the discipline that deals with natural and artificial system composed of many individuals that coordinate using decentralized control and self – organization in particular, the discipline focuses on collective behavior that results from local interaction of the individuals with each other and with their environment[3].
A high level of view of a system suggest that N agents in the swarm are cooperating to achieve purposeful behavior and achieve some goal, this apparent "collective intelligence" seems to emerge from what are often large groups of relatively simple agents, the agents use simple local rules to govern their actions and via interaction of the entire group, a type of the (self organization) emerges from the collection actions of the group[4]. The swarm framework or meta formalism present a unified scheme or approaching the problems and investigating the ways to implement swarm intelligence, it addresses how swarm of the entities must communicate and modify their behavior in response to information from other entities and their environment for there to exist the emergent self-organized behavior known as "swarm intelligence"[5]. There are important characteristics of the swarm colony:-

- Flexible:-the colony can respond to the internal perturbations and external challenges.
- Robust:-tasks are completed even of some individuals fail.
- Decentralized :-there is no control in the colony.
- Self-organized:-paths to the solution are emergent rather than predefined.

4. Behavior of bees in nature

Before illustrating bee swarm algorithm, we need to understand bee in nature, the best example is the collection and processing of the nature, the practice of which is highly organized each bee decides the nectar source by following a nest mate has already discovered a patch of flowers, each hive has a so called dance area in which the bee that have discovered nectar source dance in dance in that way trying to follow them if the bee decides to leave the hive to the nectar, this bee follows one of the bee dancers to one of the nectar areas, bees communicate through this waggle dance which contain the direction, the distance and the quality (fitness) of the flowers patch , these information helps the colony to send it's bees[6,7], Tereshko explains the main components of his model as below :-

1- food source: in order to select a food source, a forager bee evaluates several properties related with the food source such as its closeness to the hive, richness of energy, taste of its nectar, and the ease or difficulty of extracting this energy. For the simplicity, the quality of a food source can be represented by only one quantity although it depends on various parameters mentioned above.
2- Employed foragers: an employed forager is employed at specific food source which she is currently exploiting. She carries information about this specific source and shares it with other bees waiting in the hive. The information includes the distance, the direction and the profitability of the food source.
3- Unemployed foragers: a forager that looks for a food source to exploit is called unemployed. It can be either a scout who searches the environment randomly or an onlooker who tries to find a food source by means of the information given by the employed bee.

The number of scout s is about 5-10%. The exchange of information among bees is the most important occurrence in the formation of collective knowledge. While examining the entire hive it is possible to distinguish some parts that commonly exits in all hives, the most important part of the hive with respect to exchanging information is the dancing area. Communication among bees related to the quality of food sources occurs in the dancing area. The related dance is called waggle dance, since information about all the current rich sources is available to an onlooker bees are those bees that are waiting on the dance area in the hive for the information to be shared by employed bees about their food sources, and then make decision to choose food source[5]. Probably she could watch numerous dances and chooses to employ her self at most profitable source. There is greater probability of onlookers choosing more profitable source since more information is circulating about more profitable sources. Employed foragers share their information with probability which is proportional to the profitability of the food sources, and the sharing of this information through waggle dancing is longer in duration. Hence, the recruitment is proportional to profitability of a food source [6,7]. In order to better understand the basic behavior characteristics of foragers, let us examine figure 1

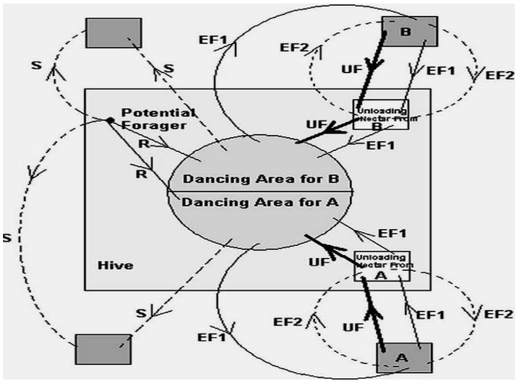


Fig. 1 bee in nature

Assume that there are two discovered food sources; A and B. at the very beginning, a potential forager will start as unemployed forager, that forager bee will have no knowledge about food sources around the nest. There are two possible options for such a bee:-

1-it can be a scout and starts searching around the nest spontaneously for food due to some internal motivation or possible external clue (S on fig 1).
2-it can be recruiting after watching the waggle dances and starts searching for food sources (R on fig 1). After finding the food source, the bee utilizes its own capability to memorize the location and then immediately starts exploiting it; hence, the bee will become an employed forager. The foraging bee takes a load of nectar from the source to the hive, unloading the nectar to food store. After Unloading the food, the bee has the following options:-
1-it might become an uncommitted follower after abandoning the food source (UF in fig1.)
2-it might dance and then recruit nest mates before returning to the same food source (EF1 in fig.1).
3-it might continue to forage at the food source without recruiting bees (EF2 in fig.1).
It is important to note that not all bees start foraging simultaneously. The experiments confirmed that new bees begin foraging at rate proportional to the difference between the eventual total number of bees and the number of the bees presently foraging [6, 7].

5-artificial bee algorithm

Artificial bee colony (ABC) algorithm is a new swarm intelligence algorithm , which was first introduced by Karabog in Erciyes university of turkey in 2005[4], and the performance of ABC was analyzed in 2007. The ABC algorithm imitates the behaviors of the real bees on searching food source and sharing the information of food sources with other bees. Since the ABC algorithm is simple in concept, easy to implement, and has fewer control parameters, it has been widely used in many fields, such as constrained optimization problems, neural networks and clustering [6, 7]. In the ABC algorithm, the colony of artificial bees is classified into there categories: employed bees, onlooker, and scouts. Employed bees are associated with particular food source which they are currently exploiting or employed at. They carry with them information about this particular source and share information to onlookers. Onlooker bees are those bees that are waiting on the dance area in the hive for the information to be shared by the employed bees about their food sources, and then make decision to choose food source. A bee carry out random search is called a scout. In the ABC algorithm, first half of the colony consists of the employed artificial bees and the second half includes the onlookers. For every food source, there is only one employed bee. In other words, the number of employed bees is equal to the number of food sources around the hive. The employed bee whose food source has been exhausted by bees becomes a scout. The position of a food source represent a possible solution to the optimization problem and the nectar amount if a food corresponds to the quality (fitness) of the associated solution represented by that food source. Onlookers are placed on the food sources by using probability based selection process. As the nectar amount of food source increases, the probability value with which the food source is preferred by onlookers increases too[7]. This is the main steps of the algorithm are given below:-

1-cycle=1
2-initiate the food source position Xi,I=1,…Sn
3-evaluate the nectar amount (fitness FIT) of food sources
4-repeat
5-employed bees phase
    For each employed bee
    Produce new food position Vi
    Calculate the value fit (i)
    Apply greedy selection mechanism
    End for
6-calculate the probability values Pi for the solution
7-onlooker bees phase
    For each onlooker bee
    Choose food source depending on Pi
    Produce new food source position Vi
    Calculate the value fit (i)
    Apply greedy selection mechanism
    End for
8-scout bee phase
    If there is an employd bee becomes scout
Then replace it with new random source position
9-memorize the best solution achieved so far
10-cycle=cycle+1
11-until cycle=maximum cycle number

In the initialization phase, the ABC algorithm generates a randomly distributed initial food source positions of SN solutions, where SN denotes the size of employed bees or onlooker bees. In the employed bees phase, each employed bee find a new food source Vi in the neighborhood of its current source Xi. The new food source is calculated using the following equation (1):

$$vij = xij + \phi ij(xij - xkj) \text{ --------------------------(1)}$$

Where $k \in$ (1, 2,..., SN) and $j \in$ (1, 2,..., D) are randomly chosen indexes, and k has to b different form i . ij is random number between [-1,1]. And then employed bee compares the new one against the current solution and memorizes the better one by means of a greedy solution mechanism. In the onlooker bees phase, each onlooker chooses a food source with a probability which related to the nectar amount (fitness) of a fork) od source shared by employed bees. The probability is calculated using the following equation (2):

$$pi = fiti \sum_{n-1}^{Sn} fiti \text{ ---------------------------------------(2)}$$

In the scout bee phase, if a food source can not be improved through a predetermined cycles called "limit", it is removed from the population and the employed bee of that food source becomes a scout. The scout bee finds a new random food source position using the equation(3) below :

$$x_i^j = x_{\min}^j + rand[1,0](x_{\max}^j - x_{\min}^j) \text{ --------------(3)}$$

Where min j and max Xj are lower and upper bounds of parameter j respectivelyv[6, 7, 8, 9].

5-crawling techniques

This section will explain some important crawling algorithms, the beginning will be with "breadth first algorithm" which is the simplest strategy, blind traversing approach, this algorithm was explored in 1994, it uses two frontiers(queues) FIFO , it is crawling links in order in which they are encountered, the problem with this algorithm is that when the frontier is full, the crawler add only link from a crawled pages , the other problem is that it traverse URL's in sequential order as these were inserted into frontier, it does not calculate the relevancy of the link , so many useless page can be crawled . to overcome this problems of the blind traversing approach[1] , a heuristic approach called out best first crawling have been studied by Cho in 1998, the following naïve best algorithm uses relevancy function to compute similarity between desired keywords and each pages, all of the above algorithms was static approach. The followed (fish and shark) algorithm are dynamic that means fetching data in time the query is issued. In fish algorithm the intuition that relevant often have relevant neighbors, thus it searches deeper under documents have been found to be relevant to the search query and stops searching in the dry areas. In shark search algorithm is development for fish, one improvement is that instead of binary (relevant/irrelevant) evaluation of document relevancy, it returns a fuzzy score (i.e a score between 0 and 1) [1, 2].

Navarat in 2006 proposed an approach to web search based on bee hive metaphor comprising of dance floor, an auditorium and dispatch rooms two simple model that describes the process of web search, in 2007 Navarat used the bee hive metaphor for online search of users predefined group of pages, authors claim that the hive determines the best routes of search and reject bad ones by experiments reported in the paper [6, 9].

6- The proposed system

The proposed system (bee crawling system) is built using VC++, and a relational database management system to store web pages, we can use any relational database like Microsoft Access , the user interface of the crawling system contains four buttons (crawl a single page, crawl a single page using bee swarm algorithm ,crawl the web from Google directory, crawl using breadth first search), the most important button is the second one (crawl a single site using bee swarm algorithm, and crawl web from Google directory) because these two buttons work depending on bee swarm algorithm , the crawl using breadth first search was putted for comparison between the traditional approach and the new proposed one see figure 2 .

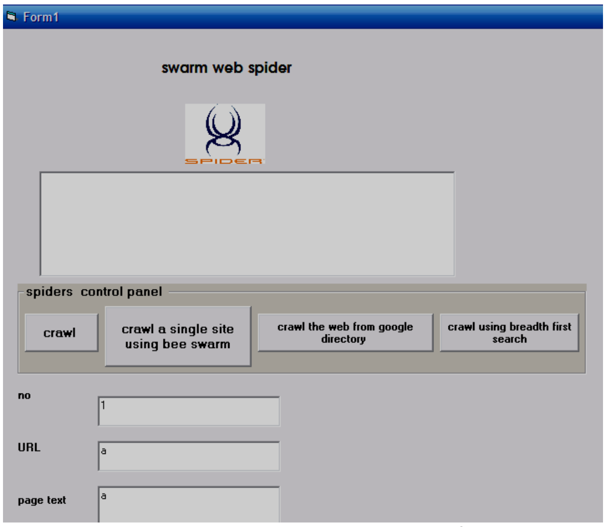


Fig. 2 crawling system main interface

The crawling algorithm depends on bee swarm intelligence algorithm as mentioned above, so the following algorithm illustrates the bee swarm crawling work:

I-start

II-initialize a crawling function by entering a seed URL to it

III-fetch the URL and save it's URL keywords and page text in the database table called(web data)
IV-find the URL's(links)to another pages
V-calculate the link relevancy using the following equation (page rank equation)
$PR(A)=(I-D)+D(PR(T1)/C(T1)+\ldots+PR(Tn)/C(Tn))$
VI-for each link in the queue
VII-if the link rank>0.7 then initialize a new function (bee) to crawl this link goto I
VIII-end for
XI-END

Where PR(A) is the page rank of page A, PR(T1) is the page rank of page T1, C(T1) is the number of outgoing links from th e page T1 , d is a damping factor in the range 0<d<1 , usually 0.85,the page rank of the web page is calculated as a sum of the page ranks of all pages linking to it (it's incoming links), divided by the number of links on each those page (outgoing links).

7-the experimental results

in order to prove the efficiency of the proposed crawler algorithm "bee crawling algorithm", we have test the system to crawl a single site using the proposed "beed algorithm" and "breadth first search", for this test we have choose for example five sites to be crawled see figure 3

www.toyota.com
www.sony.com
www.freecomputerbooks.com
www.amazon.com
www.developnew.com/books

Fig. 3 system interface after crawling Toyota web site

These sites have been crawled using the proposed bee algorithm and breadth first search and the result for crawling speed for each sit was as follows

Web site	Time of breadth search crawling	Time of bee algorithm crawling
www.toyota.com	20 sec	15sec
www.sony.com	30 sec	20 sec
www.freecomputeroks.com	40 sec	35 sec
www.amazon.com	120 sec	100 sec

Table 1: speed of breadth search crawling and proposed bee algorithm

We can understand the results in this chart

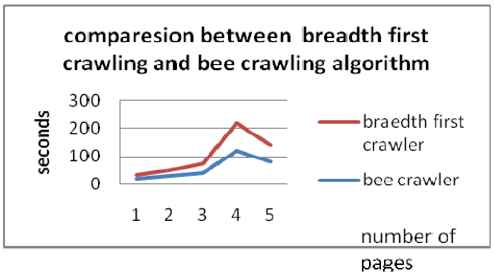


Fig. 4 chart of the results

In the obvious chart the reader can recognize that bee algorithm crawling give a good result in time than the breadth first crawling. Now, the second test is the test of crawled pages relevancy by crawling the web for specific subjects, so the user can input to the system terms like (free, computer, books, and download) with spaces between each term figure 5

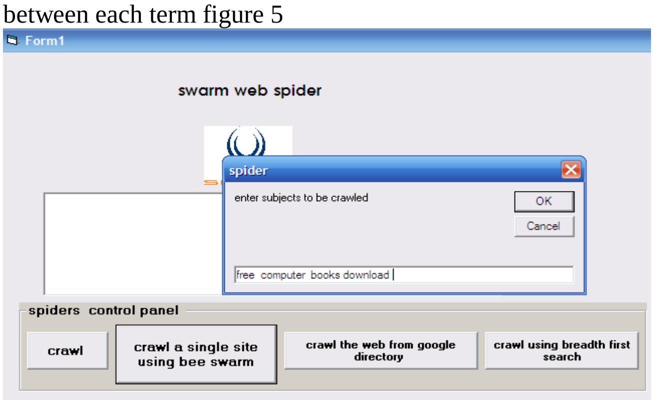


fig. 5 enter specified keywords into the system to be searched and crawled in Web

then press on "crawl the web from Google directory", this system will crawl a sample of 10 sites , now repeat the operation by using the last button "crawl using breadth first search" and also 10 sites will be crawled, see the table

Table2: the time between old and proposed method

Sites crawled using Bee algo.	Sites crawled using BF. Algo.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 1, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

1-freecomputerbooks.com 2-freebookcenter.com 3-freetechbooks.com 4-appsapps.com 5-techbooksforfee.com	1-overstock.com 2-kobobooks.com 3-amazon.com/stael-this-computer-book 4-eindiabooks.com 5-buy-ebooks.com 6-nextag.com/ebook-reader/stores.html

These sites has been crawled using breadth first search in 30 minutes while in bee algorithm in 17 minutes as the system use a function for each site, the second issue to prove the effiecny of bee swam algorithm is the relevancy of the crawled pages to the user request by measuring the relation of the frequencies of the word in these crawled sites ,this measuring has been done using a statistical tool for word mining called stat miner , so we can see the frequency of the terms and it's related to the subject of these sites which is free e-books see the table below of the frequencies of words in the crawled sites

a.txt - Notepad

File Edit Format View Help

	FREQUENCY	% SHOWN	% PROCESSED	% TOTAL	NO. CASES	% CASES
BOOKS	308	7.3%	5.0%	3.8%	1	100.0%
FREE	254	6.0%	4.1%	3.1%	1	100.0%
FREETECHBOOKS	245	5.8%	4.0%	3.0%	1	100.0%
HTTP	203	4.8%	3.3%	2.5%	1	100.0%
WWW	183	4.4%	3.0%	2.3%	1	100.0%
COMPUTER	182	4.3%	2.9%	2.2%	1	100.0%
DEVELOPNEW	145	3.4%	2.3%	1.8%	1	100.0%
ONLINE	116	2.8%	1.9%	1.4%	1	100.0%
SCIENCE	105	2.5%	1.7%	1.3%	1	100.0%
PROGRAMMING	90	2.1%	1.5%	1.1%	1	100.0%
EBOOKS	74	1.8%	1.2%	0.9%	1	100.0%
REGISTER	72	1.7%	1.2%	0.9%	1	100.0%
DOWNLOAD	69	1.6%	1.1%	0.8%	1	100.0%
HTML	69	1.6%	1.1%	0.8%	1	100.0%
LECTURE	61	1.5%	1.0%	0.8%	1	100.0%
NOTES	61	1.5%	1.0%	0.8%	1	100.0%
IGNOU	59	1.4%	1.0%	0.7%	1	100.0%
TEXTBOOKS	58	1.4%	0.9%	0.7%	1	100.0%
DESIGN	54	1.3%	0.9%	0.7%	1	100.0%
WEB	51	1.2%	0.8%	0.6%	1	100.0%
CONTACT	48	1.1%	0.8%	0.6%	1	100.0%
SEARCH	48	1.1%	0.8%	0.6%	1	100.0%
FAQ	45	1.1%	0.7%	0.6%	1	100.0%
NET	45	1.1%	0.7%	0.6%	1	100.0%
RSS	44	1.0%	0.7%	0.5%	1	100.0%
FORM	43	1.0%	0.7%	0.5%	1	100.0%
LOG	42	1.0%	0.7%	0.5%	1	100.0%
FEED	41	1.0%	0.7%	0.5%	1	100.0%
HOME	40	1.0%	0.6%	0.5%	1	100.0%
HOMEPAGE	40	1.0%	0.6%	0.5%	1	100.0%
MEMBERLIST	40	1.0%	0.6%	0.5%	1	100.0%
RESOURCES	40	1.0%	0.6%	0.5%	1	100.0%
READ	35	0.8%	0.6%	0.4%	1	100.0%
JAVA	34	0.8%	0.5%	0.4%	1	100.0%
USER	34	0.8%	0.5%	0.4%	1	100.0%

fig. 6 the words and frequencies in the crawled sites

These results can be explained by the following chart

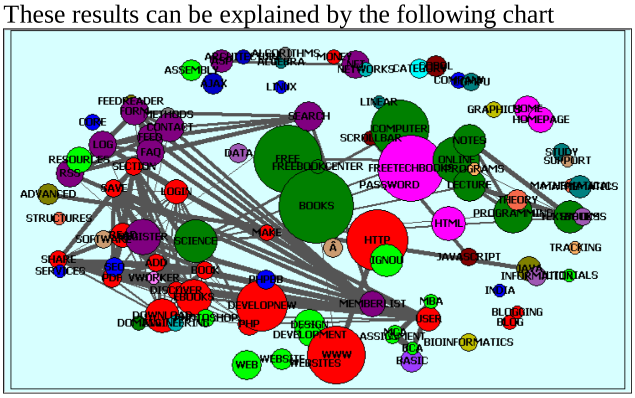


Fig. 7 the relations between keywords in sites crawled using bee algorithm

The reader can recognize from the obvious chart the close relation between terms like "free", "books", "computer", "science", "programming", "notes", and "lecture" , which appears in green color as these words coming together too much, and the relation between "http", "www", and "pdf", "websites", "make", "share", "save", so the brief result will be as the table below

Table3: the term and its frequencies

The term	Frequency
Books	308
Free	254
Free techbooks	245
Computer	182
science	105
ebooks	74
download	72

so it is a good result for "free e books" sites crawled using bee swarm algorithm, then when analyze the web sites crawled by the breadth first search algorithm from web , and analyze the relation between terms in these sites . the reader can recognijze that these sites don't represent strongly "free computer e books"web sites , see the figure

sample3.txt - Notepad

File Edit Format View Help

	FREQUENCY	% SHOWN	% PROCESSED	% TOTAL	NO. CASES	% CASES	TF • IDF
OVERSTOCK	449	7.2%	5.5%	4.4%	1	100.0%	0.0
HARDWARE	307	4.9%	3.8%	3.0%	1	100.0%	0.0
COMPUTERS	269	4.3%	3.3%	2.6%	1	100.0%	0.0
SOFTWARE	261	4.2%	3.2%	2.5%	1	100.0%	0.0
ELECTRONICS	213	3.4%	2.6%	2.1%	1	100.0%	0.0
HTTP	189	3.0%	2.3%	1.8%	1	100.0%	0.0
WWW	187	3.0%	2.3%	1.8%	1	100.0%	0.0
HTML	173	2.8%	2.1%	1.7%	1	100.0%	0.0
COMPUTER	159	2.6%	2.0%	1.5%	1	100.0%	0.0
SHIPPING	152	2.4%	1.9%	1.5%	1	100.0%	0.0
DEPT	130	2.1%	1.6%	1.3%	1	100.0%	0.0
ONLINE	111	1.8%	1.4%	1.1%	1	100.0%	0.0
CARS	95	1.5%	1.2%	0.9%	1	100.0%	0.0
SHOPPING	95	1.5%	1.2%	0.9%	1	100.0%	0.0
BUY	93	1.5%	1.1%	0.9%	1	100.0%	0.0
TRAVEL	92	1.5%	1.1%	0.9%	1	100.0%	0.0
INSURANCE	91	1.5%	1.1%	0.9%	1	100.0%	0.0
INFO	90	1.4%	1.1%	0.9%	1	100.0%	0.0
ACCESSORIES	86	1.4%	1.1%	0.8%	1	100.0%	0.0
PRICES	86	1.4%	1.1%	0.8%	1	100.0%	0.0
FACEBOOK	82	1.3%	1.0%	0.8%	1	100.0%	0.0
TWITTER	82	1.3%	1.0%	0.8%	1	100.0%	0.0
COMMUNITY	81	1.3%	1.0%	0.8%	1	100.0%	0.0
GIFTS	81	1.3%	1.0%	0.8%	1	100.0%	0.0
PINTEREST	80	1.3%	1.0%	0.8%	1	100.0%	0.0
LIST	79	1.3%	1.0%	0.8%	1	100.0%	0.0
OMAIL	79	1.3%	1.0%	0.8%	1	100.0%	0.0
AK	77	1.2%	1.0%	0.7%	1	100.0%	0.0
ENTIRE	77	1.2%	1.0%	0.7%	1	100.0%	0.0
EXCLUDES	77	1.2%	1.0%	0.7%	1	100.0%	0.0
INT	77	1.2%	1.0%	0.7%	1	100.0%	0.0
ORDER	77	1.2%	1.0%	0.7%	1	100.0%	0.0
REVIEWS	74	1.2%	0.9%	0.7%	1	100.0%	0.0
DISCOUNT	73	1.2%	0.9%	0.7%	1	100.0%	0.0
EVERYDAY	73	1.2%	0.9%	0.7%	1	100.0%	0.0

fig. 8 keyword frequency for web sites crawled by using breadth first algorithm

The reader can observe the frequency of words from the crawled sites as below

Table4: terms and its frequencies

The term	Frequency
HARDWARE	307
COMPUTERS	269
SOFTWARE	261
SHIPPING	152
ONLINE	111
SHOPPING	95
BUY	93
EBOOK	11

so the web sites retrieved by breadth first search algorithm have not the same relation to the “free computer site” like the retrieved by bee swarm algorithm , see the chart with the full analyzing for the retrieved sample below

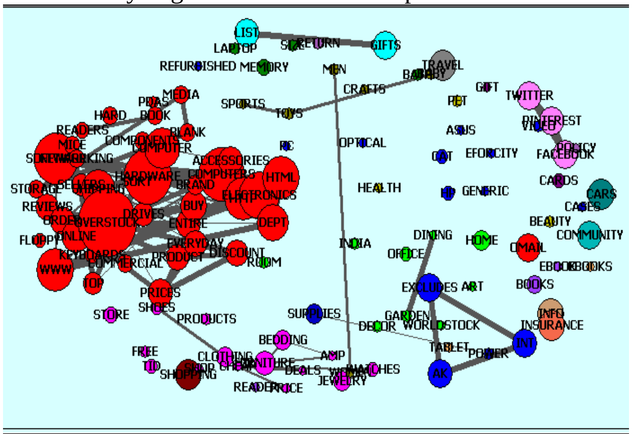


Fig.9 relations between keywords in sites which is crawled using BFS. algorithm

Lets us see the clustering analyses below

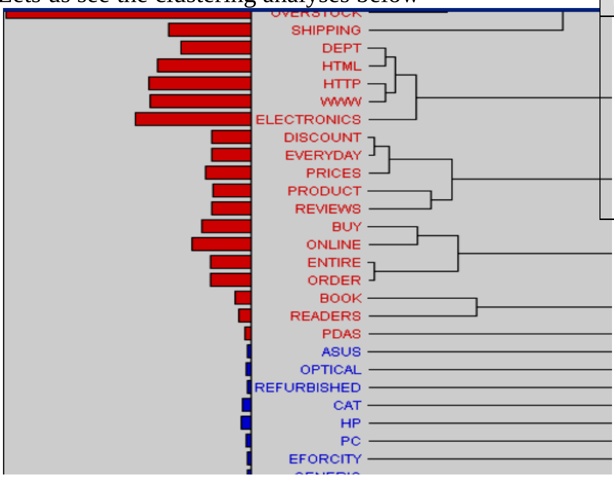


Fig. 10 clustering analysis for keywords in site crawled using breadth first algorithm

Now the second test for another terms “computer networks and books”, we try to use the system to crawl the web sites for computer networks books , because this term cause ambiguity to any searching system as it’s includes the computer network hardware and computer networking books , first we try the traditional breadth first search algorithm and second using the proposed swarm bee algorithm , so enter the term to the system as below

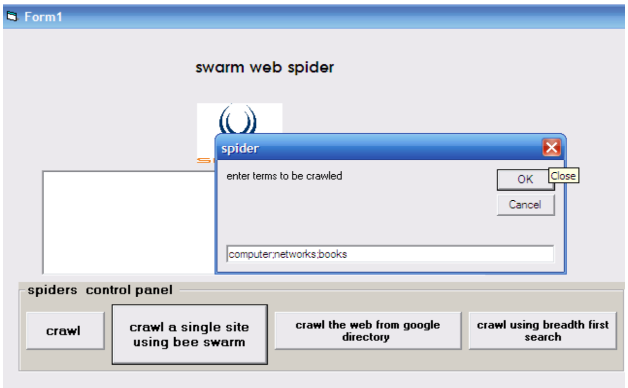


Fig. 11 enter “computer”,”network”,”books” into the system

Then press on “crawl the web from directory”, this system will crawl a sample of 10 sites, repeat the operation but using the last button “crawl using breadth first search “ and also 10 sites will be crawled

Table 5: sites crawled using the traditional method and proposed one

Sites crawled using bee algorithm	Sites crawled by breadth first algorithm
• http://compnetworking.about.com.h • http://www.amazon.com/ • http://www.target.com/ • http://freecomputerbooks.com • http://more-networking.blogspot.co • http://forums.about.com • http://linux.about.com	• http://www.cengage.com • http://www.gaebler.com • http://wps.prenhall.com • http://www.computernetworksinc.com/ • http://www.computercompany.net • http://www.cnsyn.com/ • http://www.cisco.com

To understand the relations between the keywords for the sites crawled using bee swarm algorithm see the chart below

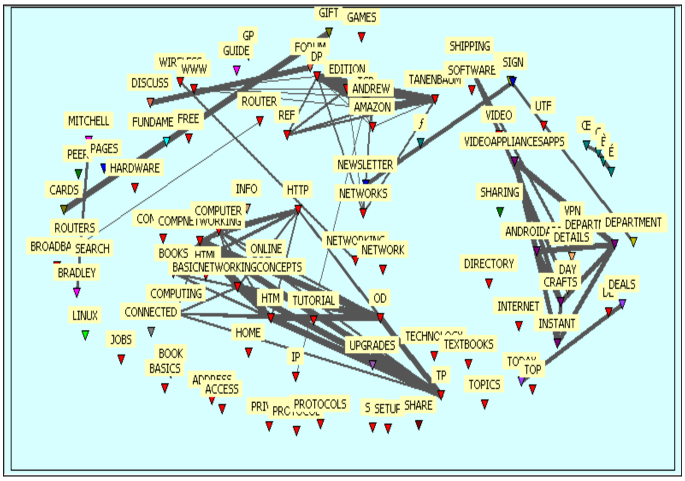


Fig. 12 the relations between words in sites crawled by bee algorithm

The reader can recognize the following frequencies for essential keywords

Table 6: keywords and it's frequencies

Keyword	frequency
Computer Networking Books Hardware	137 274 179 11

So the higher rank for computer networking books as the reader can see from the table above so the accuracy in crawling using bee algorithm is more the accuracy using breadth first search algorithm you can see also good relation between "computer networking" and books and tutorial as you see in the figure above

8-Conclusion

this research reach to a good result when using bee swarm algorithm for crawling a single web site in speed, and also we have good results in crawling the web using keywords for a group of sites in sped and relevancy of the crawled sites to the desired topics, so the a swarm intelligence can be viewed as a good improvement in web crawling area.

Refrences

[1]Sandeep Sharma,"web-crawling approachesin search ",thapar university, Patiala, India, 2008.
[2]Christopher oilstone and Mark najork, "web crawling", yahoo research, 2010.
[3]Mrco dorigo and Mauro birattari, "swarm intelligence", http://www.scholarpedia.org/article/swarm_intelligence,2007
[4]Gianni di carro, " introduction to swarm intelligence", http://www.idsia.ch/~gianni/lectures/swarm-part-1.pdf
[5]Mark Fleischer, "foundation of swarm intelligence", Institute of system research, university of Maryland ,2005.
[6]Dervis karaboga and Bahriya akay, "a comparative study of artificial bee colony algorithm", ELSEVIER, www.elsevier.com/locate/amc, 2009.
[7]Wenping Zou, Younlong Zhu, Hanning Chen, Zhu Zhu , "cooperative approaches to artificial bee colony algorithm" , international conference on computer application and system modeleing, 2010.
[8]Panta lucic and dusan teodorovic, 'computing with bees" Virginia polytechnic institute and state university ,2003.
[9]dervis karaboga and bahriye akay, "a survey: algorithms simulating bee swarm intelligence" , springer science and business media,2009.

First Author Muhamed I. Shujaa is a lecturer in the technical college of management in Iraq, has Bsc in Electrical and Computer in engineering in 1996 American university politehnical Bucharest, PhD in computer engineering 2003 at the same university and participated in many conferences and workshops about informatics inside and outside Iraq. Lecturer in informatics dept. technical college Baghdad 2004-2006. Lecturer Syrian university of technology 2006-2009.lecturer informatics dept,2010 till now. Teaching Networks,Iss,C++.

Secound Auther Ahmed bahaa aldeen is an assistance lecturer in the technical college of management in Iraq, has Bsc in computer science in 2001 from technological university and Msc. In computer science from Iraqi commission for computers and informatics in 2005, participate in many conferences and workshops about informatics inside Iraq, from 2006 till now teaching algorithms and databases in the technical college.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 1, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

Cylindrical heat exchanger trajectory planning and tracking using orthogonal functions

Hajer SAYEM[1], Naceur BENHADJ BRAIEK[2] and Hassan HAMMOURI[3]

[1] Laboratory of Advanced Systems (LSA), Polytechnic school of tunisia, BP 743, 2078 La Marsa, Tunisia

[2] Laboratory of Advanced Systems (LSA), Polytechnic school of tunisia, BP 743, 2078 La Marsa, Tunisia

[3] Laboratory of Automatic and Process Engineering (LAGEP), UMR CNRS 5007 CPE Lyon Bat. 308G 3 Bd 11 Novembre 1918, 69622, Villeurbanne cedex, France

Abstract

This paper proposes a trajectory planning and tracking approach for cylindrical heat exchanger process that is considered, under some assumptions, as a bilinear system. The proposed technique is based on orthogonal functions and especially the use of operational integration and product matrices. These operational tools allow the conversion of a bilinear differential state equation into an algebraic one depending on initial and final conditions. Arranging and solving the obtained algebraic equation lead to an open loop control law that allows the planning of a system trajectory. The parameters setting of the tracking state feedback closed loop control is yielded by considering a reference model characterizing the desired performances. A high gain observer is associated to heat exchanger process in the planning trajectory step and tracking one. A planned open loop control and a state feedback control that ensures tracking of reference trajectory were applied to the system exchanger associated to observer which is subject to noise and disturbance.

Keywords: *Feedback control, Nonlinear systems, Orthogonal functions and polynomials, Observability, heat and mass transfer.*

1. Introduction

Plan a trajectory is finding the open loop control which allows the system to reach a final state x_f set from a known initial state x_0 . Track a trajectory is to synthesize a closed loop control law that correct deviations between the real system trajectory and the planned one (so-called reference) $\Delta x = x - x_{ref} \to 0$.

Both problems were addressed in the literature in several ways: for linear systems are mentioned the technique of spline functions [1], for non linear systems notion of flatness was used [2]. However, we have proposed a new approach based on orthogonal functions as a tool of approximation to solve the planning and tracking trajectory problem for time variant linear systems [3] and bilinear system ([4],[5], [10],[11]).

In what follows we propose to apply this technique in order to plan and track a trajectory of a cylindrical heat exchanger, which is basically a six order non linear system which can be assimilated to a bilinear one by considering some practical assumptions. Moreover, a high gain observer is going to be associated with the system for state estimation.

2. Heat exchanger presentation and modeling

We present in this section the heat exchanger model with its parameters.

2.1 Process presentation and assumptions

We consider a cylindrical heat exchanger composed by three compartments shown in figure1.

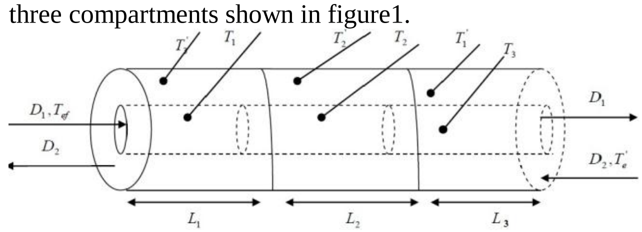


Fig. 1 Process: Three compartments heat exchanger

with:

D_i : volumetric flow

$T_e, T_e^{'}$: entry temperatures

S : section of inner cylinder

$S^{'}$: section of outer cylinder

s : heat exchange surface

$T_e, T_e^{'}$ temperatures are assumed to be homogeneous.

Also, the following assumptions are considered:

- There is no heat loss;
- There's is heat exchange through the shell of the inner cylinder under the law of Fourier: $h(T_{inner\ cylinder} - T_{outer\ cylinder}) = exchange\ flow\ (in\ watts)$ with h is the global heat transfer coefficient;
- C_p, $C_p^{'}$: constant specific heats;
- $m_i = \rho V_i = \rho S L_i$ and $m_i^{'} = \rho^{'} V_i^{'} = \rho^{'} S^{'} L_i$

2.2 Process Modeling

The modeling of the heat exchanger is based on the following thermal balance law:

$$\frac{d}{dt} m C_p T = Entry\ flow - Exit\ flow \tag{1}$$

Applying equation of thermal balance (1) for three parts I, II and III of heat exchanger.

Part I:

$$\begin{cases} \dfrac{d}{dt} m_1 C_p T_1 = \rho D_1 C_p T_e + hs(T_3^{'} - T_1) - \rho D_1 C_p T_e \\ \dfrac{d}{dt} m_1^{'} C_p^{'} T_3^{'} = \rho^{'} D_1 C_p^{'} T_2^{'} + hs(T_1 - T_3^{'}) - \rho^{'} D_2 C_p^{'} T_3^{'} \end{cases}$$

Part II:

$$\begin{cases} \dfrac{d}{dt} m_2 C_p T_2 = \rho D_1 C_p T_1 + hs(T_2^{'} - T_2) - \rho D_1 C_p T_2 \\ \dfrac{d}{dt} m_2^{'} C_p^{'} T_2^{'} = \rho^{'} D_2 C_p^{'} T_1^{'} + hs(T_2 - T_2^{'}) - \rho^{'} D_2 C_p^{'} T_2^{'} \end{cases}$$

Part III:

$$\begin{cases} \dfrac{d}{dt} m_3 C_p T_3 = \rho D_1 C_p T_2 + hs(T_1^{'} - T_3) - \rho D_1 C_p T_3 \\ \dfrac{d}{dt} m_3^{'} C_p^{'} T_1^{'} = \rho^{'} D_2 C_p^{'} T_e^{'} + hs(T_3 - T_1^{'}) - \rho^{'} D_2 C_p^{'} T_1^{'} \end{cases}$$

Let us consider the state vector of the process as $x = \begin{bmatrix} T_1 & T_2 & T_3 & T_1^{'} & T_2^{'} & T_3^{'} \end{bmatrix}^T$, a control vector $u = \begin{bmatrix} D_1 & T_e & D_2 & T_e^{'} \end{bmatrix}$ and the output $y = \begin{bmatrix} T_1 & T_3^{'} \end{bmatrix}^T$.

Note that the heat exchanger is controlled in temperature and debit.

The system evolution is described by the following state equation:

$$\begin{cases} \dot{x} = A(u)x + B(u) \\ y = Cx \end{cases} \tag{2}$$

with:

$$A(u) = \begin{pmatrix} a_{11} & 0 & 0 & 0 & 0 & a_{16} \\ a_{21} & a_{22} & 0 & 0 & a_{25} & 0 \\ 0 & a_{32} & a_{33} & a_{34} & 0 & 0 \\ 0 & 0 & a_{43} & a_{44} & 0 & 0 \\ 0 & a_{52} & 0 & a_{54} & a_{55} & 0 \\ a_{61} & 0 & 0 & 0 & a_{65} & a_{66} \end{pmatrix}$$

where:

$$a_{11} = \frac{-hs}{m_1 C_p} - \frac{\rho D_1}{m_1} \qquad a_{16} = \frac{hs}{m_1 C_p}$$

$$a_{21} = \frac{\rho D_1}{m_2} \qquad a_{22} = \frac{-hs}{m_2 C_p} - \frac{\rho D_1}{m_2}$$

$$a_{25} = \frac{hs}{m_2 C_p} \qquad a_{32} = \frac{\rho D_1}{m_3}$$

$$a_{33} = \frac{-hs}{m_3 C_p} - \frac{\rho D_1}{m_3} \qquad a_{34} = \frac{hs}{m_3 C_p}$$

$$a_{43} = \frac{hs}{m_3^{'} C_p^{'}} \qquad a_{44} = \frac{-hs}{m_3^{'} C_p^{'}} - \frac{\rho^{'} D_1}{m_3^{'}}$$

$$a_{52} = \frac{hs}{m_2^{'} C_p^{'}} \qquad a_{54} = \frac{\rho^{'} D_2}{m_2^{'}}$$

$$a_{55} = \frac{-hs}{m_2^{'} C_p^{'}} - \frac{\rho^{'} D_2}{m_2^{'}} \qquad a_{61} = \frac{hs}{m_1^{'} C_p^{'}}$$

$$a_{65} = \frac{\rho^{'} D_2}{m_1^{'}} \qquad a_{66} = \frac{-hs}{m_1^{'} C_p^{'}} - \frac{\rho^{'} D_2}{m_1^{'}}$$

$$B(u) = \begin{bmatrix} \dfrac{D_1 T_e \rho}{m_1} & 0 & 0 & \dfrac{D_2 T_e^{'} \rho^{'}}{m_3^{'}} & 0 & 0 \end{bmatrix}^T$$

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

By considering hot and cold entry temperatures $(T_e, T_e^{'})$ constant, and controlling the heat exchanger by debits D_1 and D_2 we obtain, then, a bilinear model of the exchanger; since the fourth input vector u is reduced

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 1, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

to the second order vector $u=\begin{bmatrix} D_1 & D_2 \end{bmatrix}$ and the state matrix becomes: $A(u)=A_0+A_1u$.

It is proposed to apply the trajectory planning technique based on the use of orthogonal functions developed in ([4],[6]) to the bilinear model of the heat exchanger.

3. Trajectory planning

The proposed approach to solve the problem of trajectory planning for the heat exchanger based on the use of orthogonal functions that offer the possibility of representing various systems algebraically. We first give a brief overview of the orthogonal functions.

3.1 Orthogonal functions

We consider a complete set of orthogonal functions $\Phi=\{\varphi_i(t), i\in\mathbb{N}\}$, defined on an interval $L^2([a,b])$. A projection of any function $f(t)$ in a complete space of orthogonal functions Φ is given by:

$$f(t)=\sum_{i=0}^{\infty} f_i\varphi_i(t), \quad \forall t\in[a,b] \tag{3}$$

where f_i are constant coefficients given by:

$$f_i=\frac{1}{r_i}\int_a^b \omega(x)\varphi_i(x)f(x)dx, \quad \forall i\in\mathbb{N}$$

To obtain a practical approximation of the function $f(t)$, development (3) is truncated to N order. We thus obtain:

$$f(t)\cong\sum_{i=0}^{N-1} f_i\varphi_i(t)=F_N^T\Phi_N(t) \tag{4}$$

Where: $F_N=\begin{bmatrix} f_0 & f_1 & \dots & f_{N-1}\end{bmatrix}$ is a constant coefficient vector and $\Phi_N=\begin{bmatrix}\varphi_0 & \varphi_1 & \dots & \varphi_{N-1}\end{bmatrix}^T$ is an orthogonal function composed vector.

An orthogonal functions development truncated at order N of a matrix function $A(t)=\begin{bmatrix} a_{ij}(t)\end{bmatrix}$ is given by:

$$A(t)\cong\sum_{i=0}^{N-1} A_{Ni}\varphi_i(t)$$

With $A_{Ni}\in\mathbb{R}^{n\times m}$, $i\in\{0,1,\dots,N-1\}$ are constant coefficient matrix.

Besides the approximation (4), the orthogonal functions provide useful tools such as: the operational matrix of integration, the operational matrix of product and the operational matrix of derivation, for solving differential equations.

a) Operational matrix of integration: The integral of orthogonal functions basis vector $\Phi_N(t)$ can be approximated by a constant matrix $P_N\in\mathbb{R}^{N\times N}$ which verifies: $\int_0^t \Phi_N(\tau)d\tau\cong P_N\Phi_N(t)$. Form of the matrix P_N depends on the basis of orthogonal functions chosen.

b) Operational matrix of product: The approximation of product of orthogonal basis vectors is given by the operational matrix of product M_{iN} satisfying the following relation: $\varphi_i(t)\Phi_N(t)\cong M_{iN}\Phi_N(t)$

with $M_{iN}=\begin{bmatrix} K_{0,i} & \dots & K_{N-1,i}\end{bmatrix}$ and $\forall i\in\{0,1,\dots,N-1\}$, $\varphi_i(t)\varphi_j(t)\cong K_{ij}{}^T\Phi_N(t)$

Orthogonal functions also provide product property [7] for any constant vector $V\in\mathbb{R}^n$:

$$\Phi_N(t)\Phi_N{}^T(t)V\cong M_N(V)\Phi_N(t) \tag{5}$$

Where

$$M_N(V)=\begin{bmatrix} M_{0N}(V)\vdots & M_{1N}(V)\vdots & \dots & \vdots M_{(N-1)N}(V)\end{bmatrix}$$

3.2 Proposed approach

Consider a bilinear system described by a state equation.

$$\dot{x}=Ax+\sum_{i=0}^{m} A_iu_ix+Bu \tag{6}$$

its state variables projection on an orthogonal functions basis $\Phi_N(t)$ at truncation order N gives:

$$x(t)=x_N\Phi_N(t)$$
$$u_i(t)=u_{iN}\Phi_N(t)$$
$$u(t)=u_N\Phi_N(t)$$

A state representation (6) approximation is given by:

$$\dot{x}(t)=Ax_N\Phi_N(t)+Bu_N\Phi_N(t)+\sum_{i=0}^{m} A_ix_Nu_{iN}\Phi_N(t) \tag{7}$$

integration of equation (7) between initial $t_0=0$ instant and instant t:

$$x(t)-x(0)=Ax_N\int_0^t\Phi_N(\tau)d\tau+Bu_N\int_0^t\Phi_N(\tau)d\tau + \sum_{i=0}^{m} A_ix_Nu_{iN}\int_0^t\Phi_N(\tau)d\tau \tag{8}$$

orthogonal basis projection of equation (8) and by replacing the initial state $x(0)$ (at the instant $t_0 = 0$) by its orthogonal basis projection: $x(0) = x_{N,0}\Phi_N(t)$ with $x_{N,0} = [x(0) \vdots \ 0 \vdots \ \ldots \vdots \ 0]$ and the use of operational matrix of integration P_N and product proprieties (5), one obtains:

$$x_N - x_{N,0} = Ax_N P_N \Phi_N(t) + Bu_N P_N \Phi_N(t) + \sum_{i=0}^{m} A_i x_N M_N(u_{iN}) P_N \Phi_N(t) \quad (9)$$

Using Vec operator and its main property [6]: $Vec(ABC) = (C^T \otimes A)Vec(B)$ equation (9) gives the following relation:

$$Vec(x_N) - Vec(x_{N,0}) = (P_N{}^T \otimes A)Vec(x_N) + (P_N{}^T \otimes B)Vec(u_N) + \sum_{i=0}^{m} (P_N{}^T M_N{}^T(u_{iN}) \otimes A_i)Vec(x_N)$$

thus we have:

$$Vec(x_N) = (I_{nN} - \sum_{i=0}^{m} (P_N{}^T M_N{}^T(u_{iN}) \otimes A_i)Vec(x_N) - (P_N{}^T \otimes A))^{-1}((P_N{}^T \otimes B)Vec(u_N) + Vec(x_{N,0})) \quad (10)$$

integrating equation (7) between instant t and final time ($t_{fin} = T$):

$$x(T) - x(t) = Ax_N \int_t^T \Phi_N(\tau)d\tau + Bu_N \int_t^T \Phi_N(\tau)d\tau + \sum_{i=0}^{m} A_i x_N u_{iN} \int_t^T \Phi_N(\tau)d\tau \quad (11)$$

and replace $x(T)$ by its orthogonal functions projection: $x(T) = x_{N,T}\Phi_N(T)$ and $x_{N,T} = [x(T) \vdots \ 0 \vdots \ \ldots \vdots \ 0]$ using the fact that the orthogonal basis vector at final time T verifies: $\Phi_N(T) = K_N \Phi_N(t)$

one obtains:

$$x_{N,T} - x_N = Ax_N P_N(K_N - I_N) + Bu_N P_N(K_N - I_N) + \sum_{i=0}^{m} A_i x_N M_N(u_{iN}) P_N(K_N - I_N)$$

let us pose $\Pi_N = P_N(K_N - I_N)$ and apply Vec operator, we obtain:

$$Vec(x_{N,T}) - Vec(x_N) = (\Pi_N{}^T \otimes A)Vec(x_N) + (\Pi_N{}^T \otimes B)Vec(u_N) + \sum_{i=0}^{m} (\Pi_N{}^T M_N{}^T(u_{iN}) \otimes A_i)Vec(x_N)$$

this equation yields to:

$$Vec(x_N) = (I_{nN} + \sum_{i=0}^{m} (\Pi_N{}^T M_N{}^T(u_{iN}) \otimes A_i)Vec(x_N) + (\Pi_N{}^T \otimes A))^{-1}(Vec(x_{N,T}) - (\Pi_N{}^T \otimes B)Vec(u_N)) \quad (12)$$

by equalizing (10) and (12) one obtains the following relation:

$$H_N^{-1}(Vec(x_{N,0}) + (P_N{}^T \otimes B)Vec(u_N)) = G_N^{-1}(Vec(x_{N,T}) - (\Pi_N{}^T \otimes B)Vec(u_N))$$

where:

$$H_N = H_N(u_N) = I_{nN} - R_u$$

$$R_u = (P_N{}^T \otimes A) + \sum_{i=0}^{m} (P_N{}^T M_N{}^T(u_{iN}) \otimes A_i)$$

$$G_N = I_{nN} + (\Pi_N{}^T \otimes A) + \sum_{i=0}^{m} (\Pi_N{}^T M_N{}^T(u_{iN}) \otimes A_i)$$

substituting Π_N by its expression $\Pi_N = P_N(K_N - I_N)$, thus we have: $G_N = H_N + (K_N^T \otimes I_N)R_u$

relation (12) becomes:

$$(K_N^T \otimes I_N)Z(u_N) = \Gamma(x_{N,0}, x_{N,T}) \quad (13)$$

with:

$$Z(u_N) = H_N^{-1}(u_N)((P_N{}^T \otimes B)Vec(u_N) + Vec(x_{N,0}))$$

$$\Gamma(x_{N,0}, x_{N,T}) = (K_N^T \otimes I_{Nn})Vec(x_{N,0}) + Vec(x_{N,T})$$

Open loop planning control is obtained by minimizing, with respect of u_N, the norm of the difference between the two parts of equality (13):

$$\varsigma = \left\| (K_N^T \otimes I_N)Z(u_N) - \Gamma(x_{N,0}, x_{N,T}) \right\| \quad (14)$$

The minimization can be led by using the tools provided by Matlab optimization toolbox as the "fmincon" function.

3.3 Trajectory planning of the heat exchanger

Applying the method previously developed for the bilinear model of the heat exchanger (6), using Legendre modified polynomials as a tool for the approximation [7] and a truncation order $N = 16$.

a) Numeric data :

are given below the parameter values of the heat exchanger.

$\rho = 1000\ kg/m^3$ volumetric density of the exchanger inner part.
$\rho^{'} = 1000\ kg/m^3$ volumetric density of the exchanger outer part.
$C_p = 4183\ J/kgK$ mass heat of inner part.
$C_p^{'} = 4183\ J/kgK$ mass heat of outer part.
$L = 1m$ length of heat exchanger.
$L_1 = L_2 = L_3 = L/3$ length of respectively exchanger part I, part II and part III.
$d_i = 0.03m$ inner diameter of the exchanger.
$d_e = 0.05m$ outer diameter of the exchanger.
$s_e = \dfrac{\pi d_e^2}{4}\ m^2$ section of the inner tube.
$s_e = \dfrac{\pi d_i^2}{4}\ m^2$ flow section of the fluid in the outer tube.
$m_1 = \rho s_i L_1$, $m_2 = \rho s_i L_2$, $m_3 = \rho s_i L_3$ masses of inner three parts of exchanger.
$m_1^{'} = \rho^{'} s_e L_1$, $m_2^{'} = \rho^{'} s_e L_2$, $m_3^{'} = \rho^{'} s_e L_3$ masses of outer three parts of exchanger.
$s = \pi d_i L$ exchange surface.
$T_f = T_e = 25^\circ C$ cold temperature.
$T_c = T_e^{'} = 45^\circ C$ hot temperature.
$D_f = \dfrac{4010^{-3}}{3600} m^3/s$ cold flow.
$D_c = \dfrac{16510^{-3}}{3600} m^3/s$ hot flow.
$h = 1150\ sw/m^2K$

b) Matlab simulation:

The proposed method was implemented in Matlab, for the exchanger model (2) based on previous numeric data. An open loop planning control u_p was calculated for the process reaches the final temperatures $x_f = \begin{bmatrix} 30^\circ C & 35^\circ C & 40^\circ C & 45^\circ C & 40^\circ C & 35^\circ C \end{bmatrix}^T$ with $t_{fin} = T = 20s$ from initial state $x_0 = \begin{bmatrix} 15^\circ C & 15^\circ C & 15^\circ C & 25^\circ C & 25^\circ C & 25^\circ C \end{bmatrix}^T$. The solution of equation (14) allowing the calculation of the control u_p was led by the Matlab function "fmincon" with an initial debit control $u_0 = \begin{bmatrix} D_c & 2D_c \end{bmatrix}$, a minimum level control $u_{\min} = \begin{bmatrix} D_f & D_f \end{bmatrix}$ and an upper level control $u_{\max} = 4u_0$. Modified Legendre polynomials have been used as an approximation tool with a truncation order $N = 16$.
The figure (Fig.2) show that the open loop generated trajectory allows the bilinear model of the heat exchanger to reach desired temperatures from chosen temperatures.

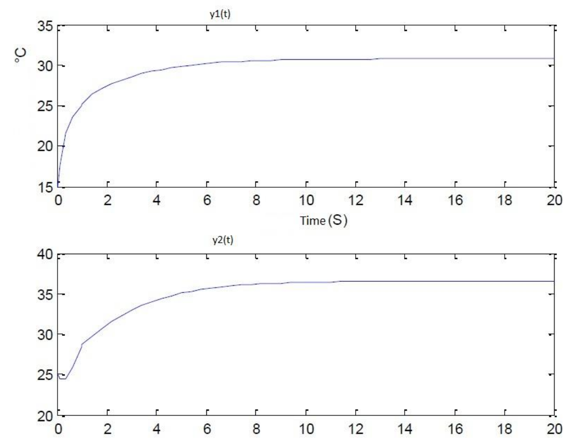


Fig. 2 Exchanger planned trajectory

4. Trajectory tracking

In this section we propose to present a method that permit to find a close loop control that ensures the planned trajectory.

4.1 Proposed approach

We consider the following difference variables between a system (6) trajectory $(x(t), u(t))$ and its planned one $(x_p(t), u_p(t))$:

$$\begin{cases} \delta x = x(t) - x_p(t) \\ \delta u = u(t) - u_p(t) \end{cases} \qquad (15)$$

the planned trajectory variables verifies:

$$\dot{x}_p = Ax_p + \sum_{i=0}^{m} A_i u_{ip} x_p + Bu_p$$

the state equation of difference system is given by:

$$\delta\dot{x} = (A + \sum_{i=0}^{m} A_i u_{ip})\delta x(t) + (B + \sum_{i=0}^{m} A_i x_p)\delta u(t) + \sum_{i=0}^{m} A_i \delta x \delta u_i \qquad (16)$$

neglecting the term $\delta x \delta u_i$ front δx and δu, the state equation (16) can be simplified to a state representation of a linear time variant system [3]:

$$\delta \dot{x} = A(t)\delta x(t) + B(t)\delta u(t) \qquad (17)$$

with: $A(t) = A + \sum_{i=0}^{m} A_i u_{ip}$

$$B(t) = B + \sum_{i=0}^{m} A_i x_p$$

Our objective is to characterize a state feedback control law $\delta u(t) = -k\delta x(t)$ that gives the controlled time variant linear system (17) the desired performances which are defined by a suitably chosen linear model:

$$\delta \dot{x} = E\delta x(t) \qquad (18)$$

the projection of time variant matrices $A(t)$, $B(t)$ and the vector $\delta x(t)$ in orthogonal functions basis is given by:

$$A(t) = \sum_{i=0}^{N-1} A_{Ni}\varphi_i(t) \qquad B(t) = \sum_{i=0}^{N-1} B_{Ni}\varphi_i(t)$$

$$\delta x(t) = \delta x_N \Phi_N(t)$$

yields the following differential equation:

$$\delta \dot{x} = \sum_{i=0}^{N-1} A_{Ni}\varphi_i(t) - k\sum_{i=0}^{N-1} B_{Ni}\varphi_i(t)\delta x_N \Phi_N(t) \qquad (19)$$

integrating the previous relation and using operational matrix of product and of integration with the use of Vec operator, we obtain:

$$\begin{aligned} Vec(\delta x_N) - Vec(\delta x_{N,0}) &= \sum_{i=0}^{N-1} (M_{iN}P_N)^T \otimes A_{Ni} \\ &\quad - k\left(\sum_{i=0}^{N-1} (M_{iN}P_N)^T \otimes B_{Ni}\right)Vec(\delta x_N) \end{aligned} \qquad (20)$$

a similar development for the reference model (18) gives:

$$Vec(\delta x_{N,r}) - Vec(\delta x_{N,0}) = (P_N^T \otimes E)Vec(\delta x_N) \qquad (21)$$

Equalization between $Vec(\delta x_N)$ obtained from (20) and $Vec(\delta x_{N,r})$ obtained from (21) provides the following linear algebraic equation whose unknown is the state feedback control gain k:

$$\phi k = \psi \qquad (22)$$

with:

$$\phi = \sum_{i=0}^{N-1} (M_{iN}P_N)^T \otimes B_{Ni}$$

$$\psi = \sum_{i=0}^{N-1} (M_{iN}P_N)^T \otimes A_{Ni} - (P_N^T \otimes E)$$

Solving the equation (22) using the least squares method leads to a state feedback control law $\delta x(t) = -k\delta u(t)$ that tracks the bilinear system (6) trajectory.

It should be noted that the development (17) to (22) can be extended to synthesize a time variant state feedback control law in which the time variant control gain $k(t)$ can be determined by a projection in an orthogonal functions base: $k(t) = \sum_{i=0}^{N-1} K_{Ni}\varphi_i(t)$.

4.2 Heat exchanger trajectory tracking

The developed method has been applied to track the planned trajectory of the heat exchanger bilinear model (6), using as reference linear model: $\delta \dot{x} = E\delta x(t)$ with $E = -I_6$.

To simulate the system provided with the obtained control law, we have introduced a perturbation on the exchanger planned trajectory at instant $t = 8s$, then we applied the closed loop control law for the disturbed trajectory. The simulation results are presented in the figure (Fig.3).

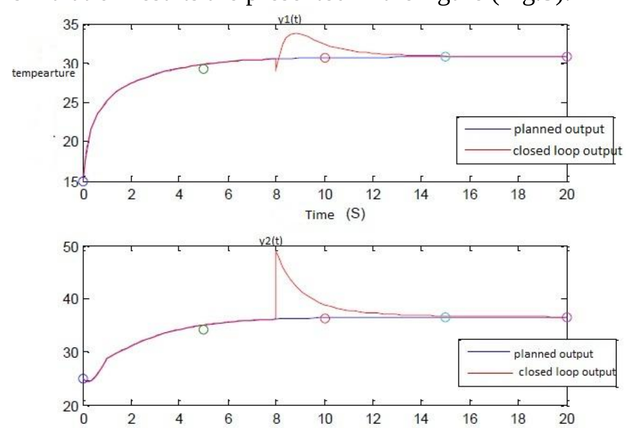


Fig. 3 Evolution of perturbed outputs of heat exchanger

Note that the closed loop system trajectory tracks the planned trajectory despite the disturbance injected into the system. The synthesized control allows the system to reach the reference trajectory (obtained in open loop) with performances defined by the choice of the linear reference model. These performances can be adjusted by changing the reference model.

5. State observer

5.1 State observer presentation

The implementation of the synthesized control and planned law needs the measure or the estimation of all the state variables. Since the temperatures $T_1^{'}$, T_2, $T_2^{'}$ and T_3 are not measurable, it is required to reconstruct them using a state observer. For this goal we propose to apply a high gain state observer ([8], [9]) described by the following equation:

$$\begin{cases} \dot{\hat{x}} = A(u)\hat{x}(t) + B(u) - S^{-1}C^T(Cx(t) - C\hat{x}(t)) \\ \dot{S}(t) = -\theta S(t) - A^T(u)S(t) - S(t)A(u) + C^T C \end{cases} \quad (23)$$

with:

θ: tuning parameter (observer gain)

$S = I$: at the first iteration.

In the next we propose to associate this state observer to the heat exchanger first in trajectory generation step, second in trajectory tracking phase.

5.2 Use of state observer in open loop step

a) Simulation without noise:

Simulating on MATLAB exchanger system (2) associated to the state observer (23) knowing that the initial state of the observer is $\bar{x}_0 = \begin{bmatrix} 10^\circ C & 10^\circ C & 10^\circ C & 20^\circ C & 20^\circ C & 20^\circ C \end{bmatrix}^T$ and the observer gain $\theta = 4$, the heat exchanger is simulated under the same conditions chosen for trajectory planning. Results are shown in figure (Fig.4).

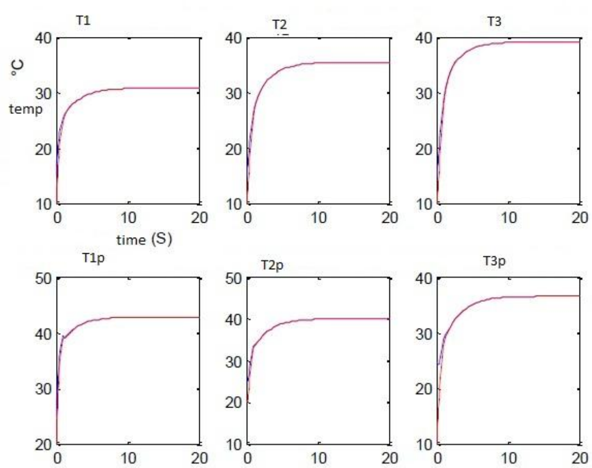


Fig. 4 Evolution of exchanger observed states without noise.

Note that starting from a state $\bar{x}_0 = \begin{bmatrix} 10^\circ C & 10^\circ C & 10^\circ C & 20^\circ C & 20^\circ C & 20^\circ C \end{bmatrix}^T$ different from the initial state $x_0 = \begin{bmatrix} 15^\circ C & 15^\circ C & 15^\circ C & 25^\circ C & 25^\circ C & 25^\circ C \end{bmatrix}^T$ chosen for the heat exchanger, the observer converges quickly to the measured output. The observer time convergence depends on the tuning parameter θ ; the higher this parameter more rapid the convergence is.

b) Simulation with noise:

Injecting a Gaussian white noise of amplitude 3% on the exchanger output and simulating the system exchanger associated with state high gain observer response, one obtains the results illustrated on the curves of figure (Fig.5).

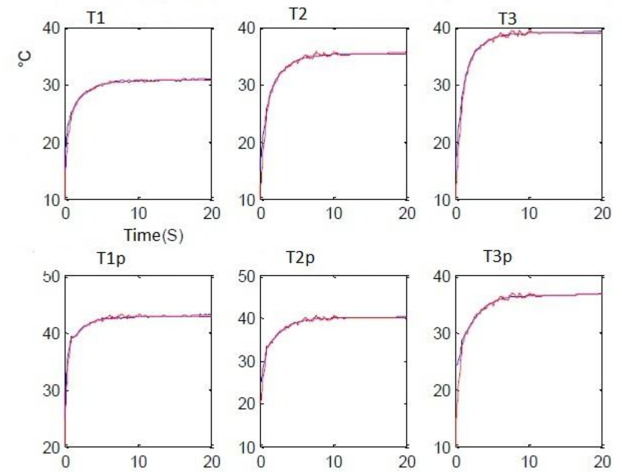


Fig. 5 Evolution of exchanger observed states with noise.

Note that despite the presence of noise on the output, the observer converges to the exchanger measured output.

5.3 Feedback control using state observer

We apply the trajectory tracking approach [11] for the system using the state observer (23) in the presence of noise by injecting at time $t = 8s$ a perturbation; the results obtained are given by figure (Fig.6).

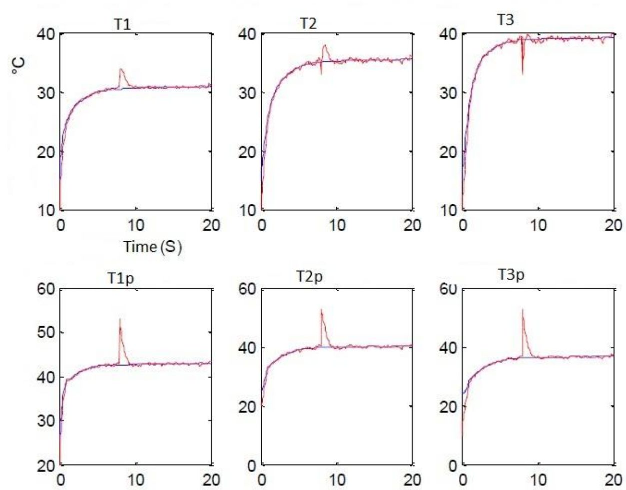


Fig. 6 Evolution of exchanger observed and controlled states with noise.

It appears that the system (exchanger+state observer + feedback control) succeed to track the planned trajectory in spite of perturbations injected.

5. Conclusion

A trajectory planning and tracking approach based on the use of orthogonal functions has been proposed and applied to a bilinear model of cylindrical heat exchanger.
This approach allowed finding an open loop planning control ensuring for system to evolve from a fixed initial state to a known final state. It should be noted that the considered control is a debit heat exchanger control that has achieved the desired temperatures for external and internal parts of the cylindrical tube.
The planned trajectory was generated in the first. Second, a tracking trajectory approach based on the use of orthogonal functions and a reference model was applied to the exchanger bilinear model. Then, a control state feedback law was synthesized allowing controlled system to stay in a neighborhood of the reference trajectory despite disturbances that may occur.
Moreover, a high gain state observer of the heat exchanger has been integrated in the control loop which has permitted the tracking of the planned trajectory despite of the non measurability of the system state variables and the noise which may occur on the process.

References

[1] M. Egerstedt. and Ph. Martin. «Trajectory Planning for Linear Control Systems with Generalized Splines». In: Proceedings of the mathematical theory of networks and systems. Padova, Italy, 1998, 102–106.
[2] M. Fliess, J. Lévine, Ph. Martin and P. Rouchon. «Design of trajectory stabilizing feedback for driftless flat systems. » In: Proc of the 3rd European Control Conf, Rome, 1995, 1882-1887.
[3] N. Benhadj Braiek, H. Sayem and H. Hammouri. « Approche de planification et suivi de trajectoires pour les systèmes non stationnaires par utilisation des fonctions orthogonales ». In: Conférence Francophone d'Automatique CIFA, Bucarest-Roumanie, Septembre 2008.
[4] H. Sayem, N. Benhadj Braiek and H. Hammouri. "Trajectory planning and tracking of nonlinear systems using orthogonal functions". In: 1er Atelier International sur les Systèmes ingénierie Conception et Application SENDA 2008, Monastir-Tunisie, Octobre 2008.
[5] H. Sayem, N. Benhadj Braiek and H. Hammouri. " Planning and tracking trajectory of bilinear systems using orthogonal functions". In: Nonlinear Dynamics And Systems Theory (ND&ST),volume 10 issue 3 pp.295-304.2010.
[6] J.W. Brewer. «Kronecker Products and matrix calculus in systems theory». In: IEEE Trans Circ .and Syst.,CAS-25, (1978).
[7] B. Ayadi and N. Benhadj Braiek. « Commande PID des systèmes linéaires à temps variant par utilisation des fonctions orthogonales ». In: Conférence Francophone d'Automatique CIFA, Bordeaux-France, 2006.
[8] J.P. Gauthier, H.Hammouri and Othman. «A simple observer for nonlinear systems-Application to bioreactors». In:IEEE Trans. Aut.contol Vol. 37,N6 june,1992.
[9] H. Hammouri. «Uniform observability and observer synthesis». In: G. Besanon (ED): Nonlinear Observers and Applications. LNCIS 363, pp.35-69.2007.
[10] H. Sayem, N. Benhadj Braiek an H. Hammouri. « Trajectory planning of bilinear systems using direct and indirect methods». In: 13th International Conference on Sciences and Techniques of Automatic control and computer engineering (STA'2012). Tunisia, December 2012.
[11] H. Sayem, N. Benhadj Braiek an H. Hammouri. « Trajectory tacking of bilinear systems using high gain observer». In: International Conference on Electrical Engineering and Software Applications ICEESA. Tunisia, March 2012.

Dr. Hajer Sayem she is graduated from National school of engineering of Monastir, Tunisia. Member of Advanced Systems Laboratory LSA Polytechnic school of Tunisia.

Pr. Naceur Benhadj Braiek he is a professor on high engineering school of Tunis and director of Laboratory of advanced systems LSA, Polytechnic school of Tunisia.

Pr. Hassan Hammouri professor on Claude Bernard University of Lyon France.

Comparative Study Based on Open Source Content Management Systems Mambo and his Fork – Joomla and Elxis

Dejan Viduka[1], Igor Lavrnic[1] and Ana Basic[2]

**[1] Department of Postgraduate Studies and International Cooperation, University Singidunum
Belgrade, 11000, Serbia**

**[2] Republic Agency for Electronic Communication
Belgrade, 11000, Serbia**

Abstract

The paper is dedicated to content management system Mambo evolution and its derivatives (Joomla i Elxis) who created a revolution in the field of web development and design. This paper explains all basic terms regarding content management system, open source, fork bases, historical development of Mambo and its derivatives (forks) as well as comparative analyzes of their characteristics. The CMS systems are described, as comparative analyzes of their results of work speed.. Finally the author made a recommendation regarding the user friendly characteristics, indicating their advantages and disadvantages.

Keywords: *Open Source, Mambo, Joomla, Elxis,* Content management system.

1. Introduction

Currently on the market are present a lot of content management systems which are available to users. Theese above mentioned products are design to be maximum user friendly with their simplicity and other advantages. Certain number of theese products are launched on the market as open source softwares under GNU licence. This licence alows increased flexibility and expandability. CMS is presenting a content management system which is covering all possible solutions allowing contents clasification, organisation, connecting and every other solution of content editing. This term can be used for manual processes of content management although it is mostly implicating on various software solutions aneabling advanced management of large number of informations. The most present form of CMS implementation is on Internet known as WCMS (Web Content Management System). One of pioneer WCMS solutions was published under the name of Mambo, achieved great suxcess due its using simpicity.

1.1 Term Open Source

Currently there are two diferent interpretations: "Free Software" and " Open Source".

The term "Free Softweare" has root at GNU project and it could be defined as follows: free softweare is mather of choice not the price.

Image 1: Official Open Source Logo

In order to understand this concept we should think in a way that "free" are like "freedom of speech" not like "free beer". Free of charge software is "users right to use the software, make a copy, distrubute, change and eventualy improve "issue. It is important to be mentioned that Open Source is not free of charge software nor only that its code is available to all users. Depending on the license under which softweare has been launched give us a wide scope of abilities of improvments and re-distribution under diferent name (fork) as well as free of charge using. Moreover the free use depends of license owners, some of them could be used only in private, and commercial use is not allowed by license owner. Finally there is a diference between Open Source and commercial software where the author is willing to make some profit, but in same time the commercial softweare could be "open code" like it is in case of PHP computer language use.

1.2 Term FORK (Fork - software development)

In software engineering, project Fork is term that occures when developers takes a copy of source code from one of software package and start an idenpendent development under different name, creating an entirely new software. Most often it happens with open source software, such as in this case with Mambo CMS, which is used as a base for creating Joomla and Elxis softwares. In case of take-over of an Open Source software and continuing with independent development, it could be done by developers without previous approval from the author,due to fact that software is published under public license (mostly GNU-GPL) which allows that.

2. Mambo CMS

Mambo is modular CMS which approach the market with a number „Core“ modules which enable its functioning. This system is present on the market since year 2000, when Miro Construct Pty Ltd started a development of Mambo CMS commercial type under leadership of Peter Lemont. Next year (2001) the company accepted double licensing, launched Mambo Site Server at Sourceforge site, under GPL license. By middle of 2002 Miro was sole developer of Mambo, contributing to security patches and bag-problems solving but not extending existing features and adding new one to this CMS. In 2002 Miro Construct Pty Ltd launched Mambo 3.0.x Open Source. Mambo Site server becomes „Mambo Open Source“ (better known as MOS). Robert Castle has been promoted to mambo Open Source project director, immediately started to create a team of developers - volunteers.

In 2003 was launched Mambo Open Source 4.0 and Miro Construct Pty Ltd officially give up the further responsibility of the developing team of Open Source Project. Furthermore Miro Construct Pty Ltd was more focused its own commercial products, while Mambo Open Source continues to work independently under Mr. Robert Castle management. Moreover Miro Construct Pty Ltd launched Mambo CMS commercial version, claiming that product do not contains the code after it was announced as an open source. In the end of 2003 Mambo Open Source 4.5 was launched, by which time almost all the code written by Miro Construct Pty Ltd disappears.

In the middle of 2004 Mamboforge.net began with work and Linux Format awarded Mambo Open Source as „Best Free Software Project“ of the year.

Linux users and developers awarded MOS as „Best Linux or Open Source Software“. After these events the name „Mambo Open Source” (known as MOS) has changed the name to „Mambo“ causing the community uneasiness about CMS who was developed together by community and Miro Construct Pty Ltd commercial solution (Mambo CMS).

Mambo version 4.5.1 was launched in September 2003 and one month later was launched“ 4.5.1 a stable“ version. Mr. Robert Castle resigned on director of project position in November 2003, and Mr. Edi Endru takeover this position couple day later. Mambo management Board has been assembled from Miro Construct Pty Ltd and Mambo development team. The Management Board has been founded to manage Mambo Project. Moreover in January 2005 Mr. Endru Edi announced a strategic partnership between Mambo and Miro International Pry Ltd. Afterwards during year 2005 it was established a mambo-foundation.org domain and Mambo Foundation.Inc as well. Mambo continues as successful Open Source Project, but later on whole team of developers left Mambo Company.

The Mambos former development team created a new company under the name “Open Source Matters“ and their new project under a name „Joomla“, and most of Mambo users started to use services of „Open Source Matters“. Later on Open Source Matters Inc. Was registered as none profit organization in New York, USA.

By the end of 2005 Joomla takeover Mambo market position and launched its first version of Mambo under name „Joomla 1.0“. In spite the fact that both CMS look almost identically, Joomla made a considerable progress which has been noticed by users which continue to follows the Joomla new solutions.

Beside Joomlas team in 2006 two Greek developers founded Elxis which was based on Mambo and Joomla as well. Elxis bring some new solutions which haven’t been recognised and supported by users, like it was in case with Joomla.

3. Joomla CMS

The Joomla solution was the most powerfull at the time as well as most awarded and used Open Source CMS system. This sotware was launched under the GNU GPL license so that any user can use and make the changes according to his needs. The advantage of this system was in simplicity in publishing, as well as changing and deleting of published contents on web pages. The reasons behind were in inbuilt text editors (WYSIWYG) and learinig to use was on level of use of MS Office Word (Open/Libre Office) packages. All issues which are connected with code, HTML or classic WEB programing was invisible for client. The market for this product was huge and through a large number of translations and varaities could explain a big popularity of this system. Basically there is nothing that you can't do with this system, with small efforts investing in work and a researching on internet, the user could find more than 1600 modules (free download) available, alowing to user tremendous opportunities for creativity.

4. Elxis CMS

Elxis is free of charge Open Source CMS offering freedom and power to do what ever you want, with no limits. Elxis created a small team of developers from Greece and Serbia. Beside development unit in project was involved and Elxis Community team providing fesh ideas, testing, fine tunning, writing documantation and translations. First three versions of Elxisa was launched under GPL license, while the last version was launched under EPL license (elxis public license). The system follows W3C standard, good security system, extremly flexible and easy to use. During development process a lot of efforts has been invested in optimisaton of search engine, which classified them in group of systems with high performance. Elxis CMS fourth generation (4.x) has been developed by full object oriented programing which was deleoped from scratch (first time solving the code issue inherited from Mambo). Only 0.01% is participation of old code from previuos version. Elxis CMS is multilingual system supporting UTF-8 coding , meaning it is able to retrive the content on any of existing language. Elxis CMS is able to translate automatically the text and keep original formating. The user needs to chuse the language and than to activate the icon translate, and in a few seconds outcomes a translated text. Elxis CMS is user frendly system available to be used very easy by new user .

5. Comparative Study

In this study are comparatively presented an info regarding 3 CMS based on following set criteria:

- Usability,
- Technical Requirements,
- User Friendly Characteristics,
- Performance,
- Certainty,
- Built In Application and
- Support.

In the table below we can be observe that all 3 software solutions support same values or in other words they are the same CMS with some improvements. Some of major improvements that have brought Joomla and Elxis are in the area of safety; performances as well as a lot of adjustments have been done in the field of search engines (SEO). Another improvement which Elxis last generation brought on the market is new license ELP- Elxis Public License, which impressed CMS supporters, but that doesn't mean that it is worse solution than Joomla. In the paper follows the test of speed and load which are more valid to view the performances of the two systems than it could be observed visually.

CMS	Mambo CMS	Elxis CMS	Joomla CMS
Usability			
WYSIWYG Editor	Yes	Yes	Yes
Version	Yes	Yes	Yes
Technical requirements			
Server	Apache Recommended	Apache Recommended	Apache Recommended
Database	MySQL Recommended	MySQL Recommended	MySQL Recommended
Operating System	Any	Any	Any
Programming Language	PHP	PHP	PHP
Shell Access	No	No	No
License	GNU-GPL	GNU-GPL / EPL	GNU-GPL
User friendly			

Friendly URIs	No	Yes	Yes
Web Robots Exclusion	Yes	Yes	Yes
Performance			
Load Balancing	No	No	No
Visitor tracking	No	Yes	Yes
Certainty			
Audit Trail	No	Yes	No
Captcha	No	Yes	Yes
Email Verification	Yes	Yes	Yes
Session Management	No	Yes	Yes
SSL Compatible	No	Yes	Free Add on for LDAP
Means of Authentication	No	Yes	No
Built-in applications			
Search Engine	No	Yes	Yes
User Contributions	Yes	Yes	Yes
Syndicated Content	Yes	Yes	Yes
Support			
Developer Community	Yes	Yes	Yes
Public Forum	Yes	Yes	Yes

Tabel 1: Comparison of three CMS (Mambo, Elxis and Joomla)

6. Elxis vs. Joomla benchmark comparison

The benchmark results of the comparing test between Joomla 2.5.4 Emper (stable) and Elxis 4.0 Nautilus (pre-alpha) will follow up. The test was implemented with Apache's ab tool, on computer which was used for local development based on MS Windows operative system. Both CMS have had same basic (default) settings, but Elxis has had 2900 published articles more than Joomla. Various caching options were tested and result shows that Elxis least four time is faster than his competitor Joomla.

6.1 Test environment

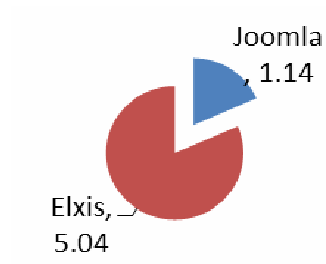
Web server: *Apache 2.2.17*
PHP version: *5.3.5*
ApacheBench: *2.3*
OS: *Windows Vista 64bit Home Premium Edition Service Pack 2*
Joomla: *2.5.4 Ember stable*
Elxis: *4.0 Nautilus rev1140 pre-alpha*

6.2 Metodology of the test

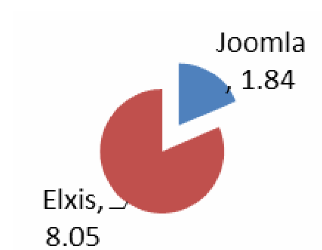
Primarily the speed testing were conducted in several different phases. The same testing was repeated twice without coming identical results. Furthermore both CMS front page performances were tested with and without turn on cash option. Moreover the content of the published articles of both CMS were tested as well. In the caching (cash) option were used following: without cash, with cash files and APC cache. Elxis APC has op-code caching based on APC caching module in caching format.

6.3 Comparison of the results

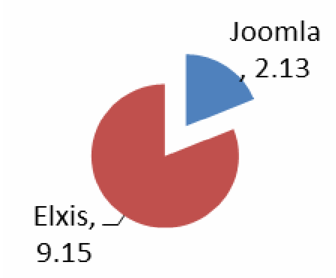
Graphs below show comparative results using the option "number of requests per second".The graph shows that Elxis is faster than Joomla even if compare Elxis without cache and Joomla best option (APC Caching). Nonetheless we should have in our mind that Elxis is still in development phase.



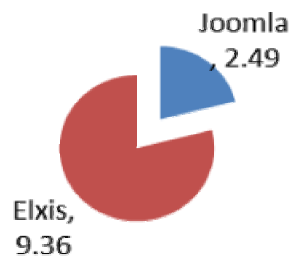
Graph 1: Result wituot cashe (FrontPage)



Graph 2: Result with file chaching (FrontPage)



Graph 3: APC result (FrontPage)



Graph 4: APC result and content testing

7. Description of the results

In the presented texts it could be observed that Elxis is faster than Joomla. Tests were conducted in the same enviroment with sole excpetion that Joomla was in the final version, against Elxis in test version, which provides more space for Elxis further improvements regarding the speed and other issues. The main reason behind why is Elxis is faster is laying in good optimisation and organisation. When the APC is turned on , Elxis reduces the number of requests sent to database located at Fronpage. In same time Elxis has the same princip of operation with contents and modules which are loaded only when they are needed . Furthermore Joomla caches all modules regardless wherthere they are published or not. One of reason why , Elxis has only two types of global extensions (components and modules) against Joomla who has 11 diferent types of extensions.

8. Conclusions

What can be observed in this work is strong influence of Mambo on development of above mentioned two CMS with great regret due Mambo ceased to exist. The paper presents detailed comparison and analysis of the three CMS. The main idea of CMS is content management providing easy and efficient way to handle it. This is an inspiration and goal of all three systems, providing easy and fast content management in order to be maximum user friendly. Joomla and Elxis have their advantages and disadvantages which is leading us to a conclusion that Joomla has excellent systematic solution, followed with a significant number of users, providing them with large number of accessories (plug in) allowing to beginners to create fabulous sites. On the other side Elxis achieved better speed which is already above described, beside that Elxis is safer and user friendly. All this is leading us to a conclusion that Joomla is more dedicated to small and medium sites, and Elxis is able to support more demanding projects as well as experienced computer scientists leading them to achieve serious results.

Acknowledgments

Using this opportunity to thank Open Source Communitty gathered around CMS systems as well as members of Elxis development team especially to Mr Ioannis Sannos and Mr Ivan Trebjesanin for their generous support during the writing of this paper.

References

[1] S.K. Patel, V.R. Rathod, and N.A. Patel, Open Source CMS Selection – A Mystery, International Journal on Computer Science and Engineering, 2010, pp. 0975-3397.
[2] S.K. Patel, V.R. Rathod and J.B. Prajapati, Performance Analysis of Content Management Systems – Joomla, Drupal and WordPress, International Journal of Computer Applications, 2011, pp. 0975-8887.
[3] B. Bonfield and L.S. Quinn, Comparing Open Source CMSes: Joomla, Drupal and Plone, Idealwore idealwore.org, 2007.
[4] N. Singhal, T. Mohan and S. Sarkar, A Comparative Study Based on Open Source Content Management Systems, Vol. 1, No.4 267-276, pp. 0976-5166.
[5] Wikipedia http://en.wikipedia.org/wiki/Fork_%28 software_development%29
[6] Wikipedia http://en.wikipedia.org/wiki/Content_management_system
[7] Official Elxis CMS, http://www.elxis.org
[8] Official Joomla CMS, http://www.joomla.org

MSc Dejan Viduka was born in Osijek, Croatia in July 30, 1980. Currently he is pursuing Ph. D. area of Technical Science at Singidunum University Belgrade, Serbia. Having 15 years of experience at IT sector (hardware, software, SEO, Internet, e-business, e-marketing, operating systems, CMS systems and Open Source systems) and extensive experience in the development of Open Source projects, and is also a member of Elxis Community and provides expert assistance to customers from Serbia. As an IT expert, he worked on many projects for well-known clients from Serbia and from abroad.

MSc Igor Lavrnic was born in Vinkovci, Croatia, on November 9, 1969, currently living in Belgrade, Serbia. He obtained his B.E. degree in Production Management from Technical Faculty in Zrenjanin, University of Novi Sad, Serbia in 2006. He received his Masters degree in Business Communication Science from Technical Faculty in Zrenjanin, Novi Sad University, Serbia in 2010. Working in banking sector since 2003 on senior management positions. Currently she is pursuing Ph. D. in the area of Business Continuity and Disaster Recovery at Department of Technical Science of Singidunum University Belgrade, Serbia.

MSc Ana Basic was born in Belgrade, Republic of Serbia, on April 22, 1981. She received the Master of Science Degree in engineering from the Faculty of Electrical Engineering, University of Belgrade, in 2006. Since 2006 she is working in Republic Agency for electronic communication, in the Department of regulation. Since 2010 she is licensed designer for telecommunication networks and systems of the Serbian Chamber of Engineers. Currently attending PhD studies at Singidunum University in Belgrade, department of Technical Science.

An Efficient and Secure Handover Protocol for IEEE 802.16m Networks

Heba K. Aslan

Informatics Dept., Electronics Research Institute
Cairo, Egypt

Abstract

Mobile WiMAX (Worldwide Interoperability Microwave Access) requires the re-authentication of mobile stations as they change from one base station to another. IEEE 802.16e uses the Extensible Authentication Protocol (EAP) for authentication and key management. This requires about 1000 ms, therefore, it could not support mobile WiMAX applications such as video conference. In the present paper, we propose a protocol that aims to overcome the inter-domain handover problem. The proposed protocol uses the EAP protocol for authentication and key distribution. The proposed protocol is based on the use of hash functions and the Diffie-Hellman protocol to distribute the keys and to avoid the domino effect. The proposed protocol is analyzed using the BAN logic to ensure that it achieves the goals of authentication and key distribution. Furthermore, the proposed protocol is compared with other handover protocols. The comparison shows that proposed protocol outperforms the other protocols.

Keywords: Wimax Security, Authentication, Handover, 802.16m.

1. Introduction

Mobile WiMAX (Worldwide Interoperability Microwave Access) is a wireless networking system based on the IEEE 802.16e standard [1]. IEEE 802.16e aims to amend 802.16-2004 [2] and to provide mobility. Mobility means that one Mobile Station (MS) can change from one Base Station to another. This requires the re-authentication of the MS. Recently, IEEE issued a new version, IEEE 802.16m [3], as an advanced air interface to meet requirements of the fourth generation (4G) systems. Before accessing the network, mutual authentication between MS and the network must be performed. IEEE 802.16m uses Extensible Authentication Protocol (EAP) for authentication and key management as well as RSA-based authentication protocol. Due to its flexibility, EAP has become the de facto authentication method for 802.16m [4 and 5]. Full EAP authentication latency requires about 1000 ms [6]. Therefore, it could not support mobile WiMAX applications such as video conference or streaming data. It has to be noted that according to latency guidelines identified by WiMAX forum, the latency should not exceed 50 ms in video conferencing or 100 ms in streaming data [7]. Therefore, executing a complete re-authentication in case of handover from one BS to another is not an ideal solution.

Solutions of the handover problem could be classified into two categories: pre-authentication and re-authentication. In pre-authentication handover [8, 9 and 10], MSs can pre-authenticate to the most likely BSs in order to reduce the handover authentication time. Although this solution solves the latency problem, it requires a large computation and communication overheads. In literature, many solutions have been proposed to solve the re-authentication problem. These solutions are divided into two categories [11]: public key-based schemes and symmetric key-based schemes. The public key-based schemes suffer from a high computation cost. On the other hand, symmetric key-based schemes are divided into three categories: key hierarchy, credential tickets and Security Context Transfer (SCT) mechanisms. Key hierarchy scheme is based on using key hierarchy designed especially for handover purposes. This solution reduces the handover latency but not to the extent that could be suitable for real time applications. On the other hand, some credential tickets schemes are based on incorporating a third party to securely distribute keys (such as Kerberos protocol). Other credential tickets schemes are based on using public key cryptography which results in a high computation cost. In SCT mechanisms, the security context (such as negotiated key) is transferred to a target BS using Context Transfer Protocol (CXTP). These schemes suffer from the domino effect problem in which the compromise of one BS will result in compromising other BS's that share the same security context. All the previous schemes solve the problem for intra-domain handover, where all BS's are in the same area. For inter-domain handover, the MS must perform the full EAP authentication protocol which leads to an impractical latency for mobile WiMAX applications.

In the present paper, a protocol that aims to overcome the inter-domain handover problem is proposed. The proposed protocol uses the Extensible Authentication Protocol (EAP) for authentication and key distribution. It is based on the use of hash functions and the Diffie-Hellman protocol to distribute the keys between mobile stations and base stations. In order to avoid the domino effect, the Diffie-Hellman components are distributed instead of the authentication key itself. The proposed protocol is analyzed using the BAN logic to ensure that it achieves the goals of authentication and key distribution. The analysis shows that the proposed protocol achieves its goals without bugs or redundancies. Furthermore, the proposed protocol is compared with other handover protocols. The comparison shows that the proposed protocol outperforms the other protocols. The paper is organized as follows: in Section 2, a survey of related work is detailed. Then, a description of the proposed protocol is given in Section 3. Next, logical analysis of the proposed protocol is illustrated in Section 4. Then, a comparison of the proposed protocol with other handover authentication protocols is presented in Section 5. Finally, the paper concludes in Section 6.

2. Related Work

IEEE 802.16e standard specifies three types of handover: Hard Handover (HHO), Macro Diversity Handover (MDHO) and Fast Base Station Switching (FBSS) [4]. In HHO, the MB communicates only with one BS, therefore, before connecting to a new BS, all connections with the old BS must be broken. This type of handover is simple but it is characterized by its large latency. Both MDHO and FBSS are known as soft handover where the MS could communicate with more than one BS at a time. Both the MS and BS maintain a list of BSs involved in the handover procedure. The solution of the handover problem could be classified into two categories: pre-authentication and re-authentication. In pre-authentication handover [8, 9 and 10], MSs can pre-authenticate to the most likely BSs in order to reduce the handover authentication time. Although this solution solves the latency problem, it requires a large computation and communication overheads which results from the exchange of unnecessary keys between the MS and the BSs that the MS never roams to.

In literature, many solutions have been proposed to solve the re-authentication problem. These solutions are divided into two categories [11]: public key-based schemes and symmetric key-based schemes. The public key-based schemes [12] are divided into: schemes based on ID-based cryptography [13], schemes based on blind signature schemes [14] and schemes based on capabilities [15]. These schemes suffer from a high computation cost. On the other hand, symmetric key-based schemes are divided into three categories: key hierarchy, credential tickets and Security Context Transfer (SCT) mechanisms. Key hierarchy scheme is based on using key hierarchy designed especially for handover purposes [16 and 17]. For example, the HandOver KEYing (HOKEY) [17] designed by Internet Engineer Task Force (IETF) uses EAP initiate/finish re-authentication exchange to derive a new re-authentication master key. This solution reduces the handover latency but not to the extent that could be suitable for real time applications. On the other hand, credential tickets schemes are based on incorporating a third party to securely distribute keys (such as Kerberos protocol) [18 and 19]. These schemes require a trusted third party to issue and verify the tickets. Moreover, they may require the modification of EAP protocol. In SCT mechanisms [20, 21, 22 and 23], the security context (such as negotiated key) is transferred to a target BS using Context Transfer Protocol (CXTP). Consequently, MS does not need to contact EAP server in order to perform handover. These schemes suffer from the domino effect problem in which compromising of one BS will result in compromising other BS's that share the same security context. All the previous schemes solve the problem for intra-domain handover, where all BS's are in the same area. For inter-domain handover, the MS must perform the full EAP authentication protocol which leads to an impractical latency for mobile WiMAX applications. In the next section, a description of the proposed protocol is detailed.

3. Inter-Domain Handover Authentication Protocol for IEEE 802.16m Networks

The proposed protocol consists of three phases: the initial phase, the intra-domain handover authentication phase and the inter-domain handover authentication phase. Fig. 1 shows an example of IEEE 802.16m network. In this network, the Access Service Network (ASN) consists of ASN Gateway (ASN-GW) and BS. An ASN-GW controls several BSs and has the role of forwarding authentication messages between the MS and the Authentication, Authorization and Accounting (AAA) server [11]. A BS provides WiMAX radio access for the MSs authenticated by the AAA server. In this paper, it is assumed that AAA server and all ASN-GW maintain trusted relations and have established secure connections. Also, each ASN-GW and BSs served by this ASN-GW maintain trusted relations and have established secure connections. In addition, an assumption is made that all entities share two constant values: a and p which will be used to calculate the Diffie-Hellman keys. Furthermore, a list of all ASN-GW identities and BS identities are previously delivered to MS. In the following subsections, description of the proposed protocol is detailed.

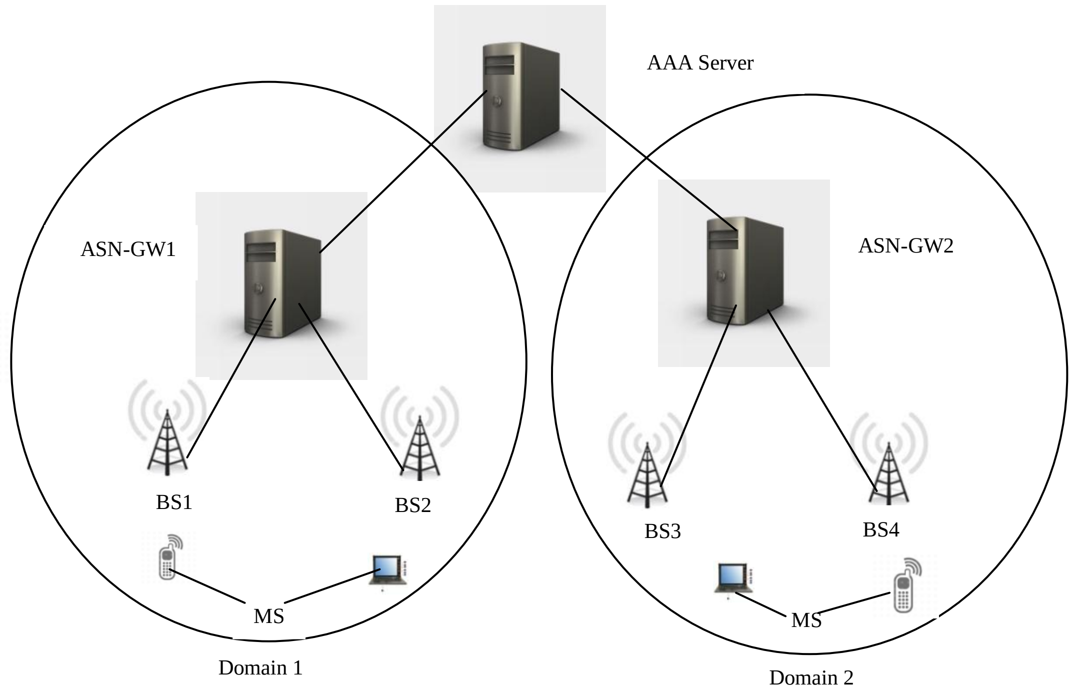


Fig. 1 IEEE 802.16m network architecture.

3.1 The Initial Phase

When an MS first accesses the IEEE802.16m network, it performs the initial authentication as shown in Fig.2. First, a full EAP authentication is executed. After a successful authentication, both AAA server and MS share a Master Session Key (MSK) which is unique for each mobile station. Then, both MS and AAA server calculates a hash value of MSK and the identity of ASN-GW1 (ASN1ID), and AAA server forwards the hashed value H_1 = H(MSK, ASN1ID) to ASN-GW1. Then, both MS and ASN-GW1 compute a Temporary CMAC Key (TCK) using H_1 and the identity of BS1 (BS1ID). Finally, ASN-GW1 forwards the temporary key TCK = H(H(MSK,ASN1ID), BS1ID) to BS1. Next, BS1 sends to the MS a message (MSG-1) containing: a timestamp T_{BS1}, the identity of BS1 (BS1ID) and its Diffie-Hellman component $a^x \bmod p$, where x is the Diffie-Hellman secret of BS1. Each entity stores and maintains the secrecy of its Diffie-Hellman component. To maintain message authenticity, a CMAC value of the abovementioned message is appended. After receiving MSG-1, MS verifies the freshness of the message using T_{BS1}, then uses TCK to validate the CMAC value. If the CMAC value is valid, MS sends to the BS1 a message (MSG-2) containing: a timestamp T_{MS}, its identity (MSID), its Diffie-Hellman component $a^y \bmod p$, where y is the Diffie-Hellman secret of MS. Again, a CMAC value of the abovementioned message is appended. Then, it calculates the Authentication Key (AK), where AK = a^{xy} mod p. After receiving MSG-2, BS1 verifies the freshness of the message using T_{MS}, then uses TCK to validate the CMAC value. If the CMAC value is valid, it calculates AK = a^{xy} mod p. Then, both BS1 and MS derive the Traffic Encryption Keys (TEK) and CMAC keys as stated in [24]. The use of Diffie-Hellman method preserves the confidentiality of the derived keys which means that only MS and BS1 know the calculated key.

3.2 The Intra-Domain Handover Authentication Phase

When an MS moves from one base station to another in the same region, it sends to ASN-GW1 a request containing its identity and the identity of the target base station, for example BS2 and appends the CMAC value of the message using TCK. The CMAC value is used to avoid man in the middle and denial of service attacks. Next, both ASN-GW1 and MS compute another temporary CMAC key (TCK') using the hash function of H_1 (hash value of MSK and ASN-GW1's identity) and the identity of BS2 (BS2ID). Next, ASN-GW1 forwards the temporary key TCK' to BS2. Next, BS2 sends to the MS a message (MSG-3) containing: a timestamp T_{BS2}, the identity of BS2 (BS1ID) and its Diffie-Hellman component $a^z \bmod p$, where z is the Diffie-Hellman secret of BS2. To maintain message authenticity, a CMAC value of the abovementioned message is appended. After receiving MSG-3, MS verifies the freshness of the message using T_{BS2}, then uses TCK' to validate the CMAC value. If the CMAC value is valid, MS sends to BS2 a message

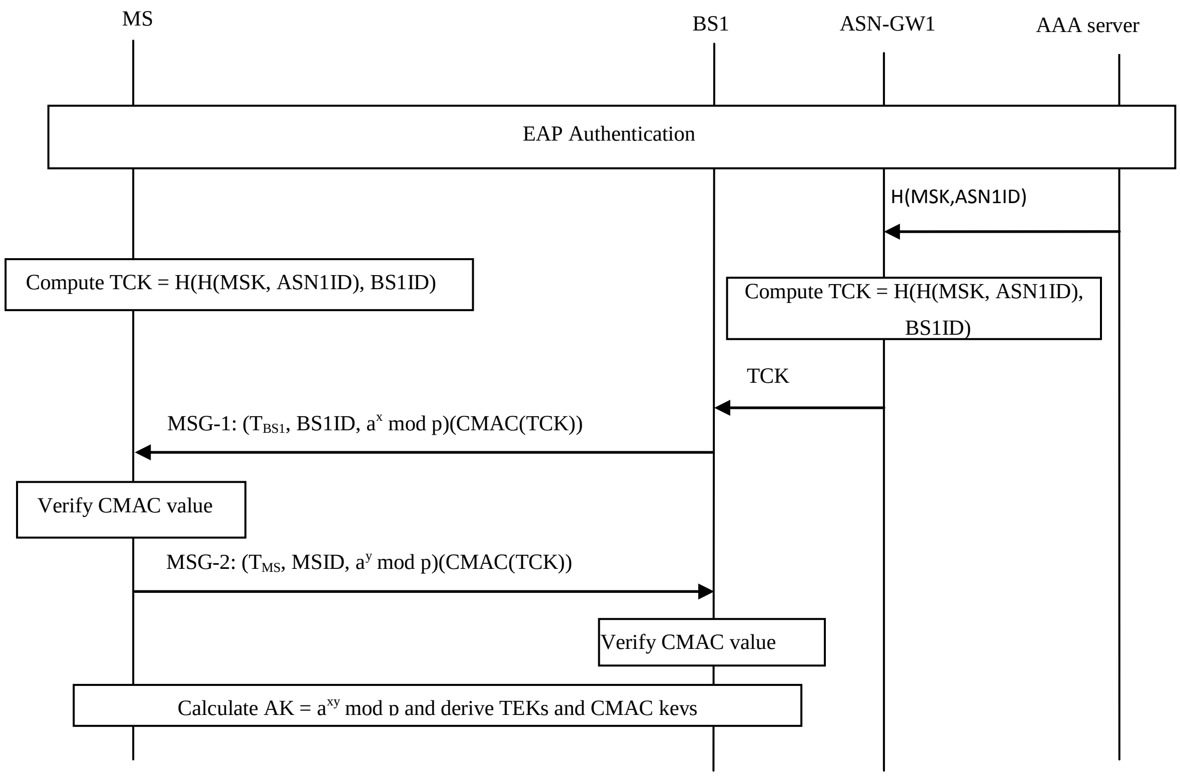


Fig. 2 Initial phase.

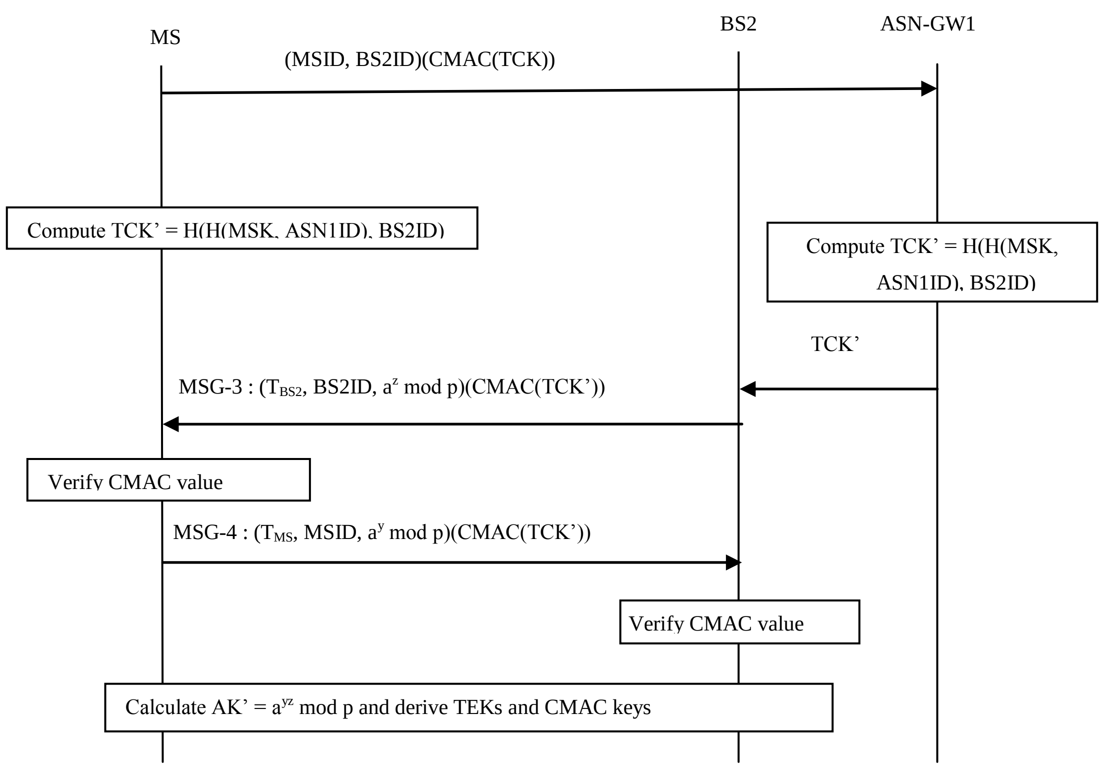


Fig. 3 Intra-domain handover phase.

(MSG-4) containing: a timestamp T_{MS}, its identity (MSID), its Diffie-Hellman component $a^y \bmod p$. Again, a CMAC value of the abovementioned message is appended. Then, it calculates the Authentication Key (AK'), where AK' = a^{yz} mod p. After receiving MSG-4, BS2 verifies the freshness of the message using T_{MS}, then uses TCK' to validate the CMAC value. If the CMAC value is valid, it calculates AK' = a^{yz} mod p. Then, both BS1 and MS derive the traffic encryption keys and CMAC keys. The abovementioned steps are illustrated in Fig. 3.

3.3 The Inter-Domain Handover Authentication Phase

When an MS moves from one base station to another in a different region, it sends to AAA server a request containing its identity and the identity of the target base station, for example BS3 and appends the CMAC value of the message using MSK. Then, both MS and AAA server calculates a hash value of MSK and the identity of ASN-GW2 (ASN2ID), and AAA server forwards the hashed value H_2 = H(MSK, ASN2ID) to ASN-GW2. Then, both MS and ASN-GW2 compute a Temporary CMAC Key (TCK'') using H_2 and the identity of BS3 (BS3ID). Finally, ASN-GW2 forwards the temporary key TCK'' = H(H(MSK,ASN2ID), BS3ID) to BS3. Next, BS3 sends to the MS a message (MSG-5) containing: a timestamp T_{BS3}, the identity of BS3 (BS3ID) and its Diffie-Hellman component $a^m \bmod p$, where m is the Diffie-Hellman secret of BS3. To maintain message authenticity, a CMAC value of the abovementioned message is appended. After receiving MSG-5, MS verifies the freshness of the message using T_{BS3}, then uses TCK'' to validate the CMAC value. If the CMAC value is valid, MS sends to the BS3 a message (MSG-6) containing: a timestamp T_{MS}, its identity (MSID), its Diffie-Hellman component $a^y \bmod p$, where y is the Diffie-Hellman secret of MS. Again, a CMAC value of the abovementioned message is appended. Then, it calculates the Authentication Key (AK''), where AK'' = a^{ym} mod p. After receiving MSG-6, BS3 verifies the freshness of the message using T_{MS}, then uses TCK'' to validate the CMAC value. If the CMAC value is valid, it calculates AK'' = a^{ym} mod p. Then, both BS3 and MS derive the Traffic Encryption Keys (TEK) and CMAC keys. The abovementioned steps are illustrated in Fig. 4. In the next section, logical analysis of the proposed protocol will be presented.

4. Logical Analysis of the Proposed Handover Protocol

BAN logic [25] is a set of rules for defining and analyzing information exchange protocols. Specifically, BAN logic helps its users to determine whether exchanged information is trustworthy, secured against eavesdropping, or both. For a successful verification of the protocol, the belief state of communicating parties should satisfy the protocol goals. The goals of the proposed protocol are: the BS and the MS believe that they share a common key and also each of them should believe that the other participant also believes in the same key.

First, the basic rules of the BAN logic are listed below:

- The interpretation rule

$$\frac{P \mid\equiv (Q \mid\sim (X, Y))}{P \mid\equiv (Q \mid\sim X), P \mid\equiv (Q \mid\sim Y)}$$

  The above rule means that if P believes that Q once said a message containing both X and Y, therefore it believes that Q once said each statement separately.

- Message Meaning Rule

$$\frac{P \equiv P \overset{Y}{\Leftrightarrow} Q, P \triangleleft < X > Y}{P \mid\equiv Q \mid\sim X}, P \neq Q$$

  This means that if P believes Y is a shared secret between it and Q, and P sees a message X combined with Y, this implies that P believes that Q once said X.

- Nonce Verification Rule

$$\frac{P \mid\equiv \#(X), P \mid\equiv Q \sim X}{P \mid\equiv Q \mid\equiv X}$$

  The above rule means that if P believes that X is a recent message and Q once said X, therefore it believes that Q believes in X.

- Jurisdiction Rule

$$\frac{P \mid\equiv Q \Rightarrow X, P \mid\equiv Q \mid\equiv X}{P \mid\equiv X}$$

  This rule means that if P believes that Q has jurisdiction over X, and P believes that Q believes in X, then P believes in X.

- Freshness Rule

$$\frac{P \mid\equiv \#(X, Y)}{P \mid\equiv \#(X)}$$

  The above rule means that if P believes in the freshness of X and Y, therefore it believes in the freshness of each statement separately.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 1, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

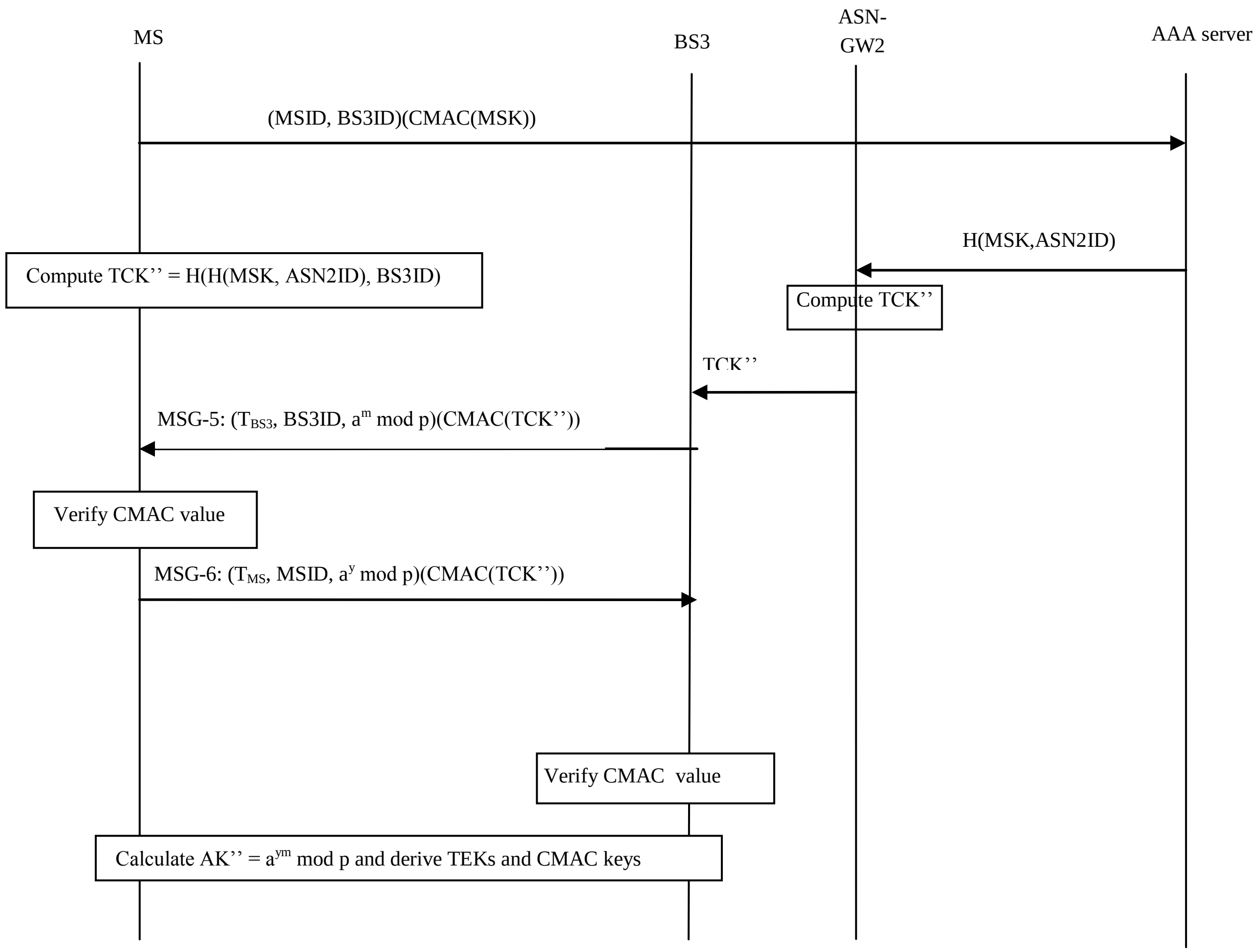


Fig. 4 Inter-domain handover phase.

- Diffie-Hellman rule

$$\frac{P \models P \underset{DH}{\overset{x}{\mapsto}} P, P \models \overset{a^{y} \bmod p}{\mapsto} Q}{P \models P \overset{a^{xy} \bmod p}{\leftrightarrow} Q}$$

This rule was added to the list of rules in order to complete the logical analysis of the proposed protocol. This rule means that if P believes that x is its Deffie-Hellman secret and it believes that a^{y}mod p is the Diffie-Hellman component of Q, therefore it believes that a^{xy}mod p is a shared symmetric key between it and Q.

The analysis is undertaken for the initial phase only, for intra-domain and inter-domain handover authentication phases, the same analysis could be carried out. Since, the authentication between the MS and the network is performed in the EAP protocol, therefore, the authentication is considered completed between the BS1 and the MS if the following goals are achieved:

Goal 1: $BS1 \models MS \overset{a^{xy} \bmod p}{\leftrightarrow} BS1$

Goal 2: $MS \models MS \overset{a^{xy} \bmod p}{\leftrightarrow} BS1$

In order to complete the analysis, the following assumptions are made:

$$BS1 \models MS \overset{TCK}{\Leftrightarrow} BS1 \quad (1)$$

$$MS \mid\equiv MS \overset{TCK}{\Leftrightarrow} BS1 \tag{2}$$

$$BS1 \mid\equiv \# T_{MS} \tag{3}$$

$$MS \mid\equiv \# T_{BS1} \tag{4}$$

$$BS1 \mid\equiv \overset{x}{\underset{DH}{\mapsto}} BS1 \tag{5}$$

$$MS \mid\equiv \overset{y}{\underset{DH}{\mapsto}} MS \tag{6}$$

$$MS \mid\equiv BS1 \Rightarrow a^{x} \bmod p \tag{7}$$

$$BS1 \mid\equiv MS \Rightarrow a^{y} \bmod p \tag{8}$$

Eqs. (1 and 2) indicate that both MS and BS1 believe that TCK is a shared secret between them. Then, Eqs. (3 and 4) indicate that both BS1 and MS believe in the freshness of T_{MS} and T_{BS1} respectively. Eqs. (5 and 6) indicate that each entity believes in its Diffie-Hellman secret. Finally, Eqs. (7 and 8) indicate that each entity believes that the other entity has jurisdiction over its Diffie-Hellman component. After making the assumptions, the messages transferred in the initial phase are transformed into logical formulas. Finally, the basic rules of the BAN logic will be applied to the logical formulas. Following is the transformation of the proposed protocol into logical formulas:

$$\text{MSG-1:} \quad BS1 \rightarrow MS : < T_{BS1}, \overset{a^{x} \bmod p}{\mapsto} BS1 > TCK \tag{9}$$

$$\text{MSG-2:} \quad MS \rightarrow BS1 : <\# T_{MS}, \overset{a^{y} \bmod p}{\mapsto} MS > TCK \tag{10}$$

The analysis of the protocol can now be performed. By applying message meaning rule to Eq. 9 and using Eq. 2, the following can be deduced:

$$MS \mid\equiv BS1 \mid\sim (T_{BS1}, \overset{a^{x} \bmod p}{\mapsto} BS1)$$

But, MS believes in the freshness of T_{BS1} (Eq.(4)). Thus, applying nonce verification rule, the following is obtained:

$$MS \mid\equiv BS1 \mid\equiv \overset{a^{x} \bmod p}{\mapsto} BS1$$

Then, by applying jurisdiction rule using Eq. (7), the following is obtained:

$$MS \mid\equiv BS1 \mid\equiv \overset{a^{x} \bmod p}{\mapsto} BS1$$

From Equation (6) and by applying the Diffie-Hellman rule, the following is obtained:

$$MS \mid\equiv MS \overset{a^{xy} \bmod p}{\leftrightarrow} BS1 \tag{11}$$

Similarly, for the analysis of the second message of the protocol, by applying message meaning rule to Eq. 10 and using Eq. 1, the following can be deduced the following:

$$BS1 \mid\equiv MS \mid\sim (T_{MS}, \overset{a^{y} \bmod p}{\mapsto} MS)$$

But, BS1 believes in the freshness of T_{MS} (Eq.(3)). Thus, applying nonce verification rule, the following is obtained:

$$BS1 \mid\equiv MS \mid\equiv \overset{a^{y} \bmod p}{\mapsto} MS$$

Then, by applying jurisdiction rule using Eq. (8), the following is derived:

$$BS1 \mid\equiv MS \mid\equiv \overset{a^{y} \bmod p}{\mapsto} MS$$

From Equation (5) and by applying the Diffie-Hellman rule, the following is obtained:

$$BS1 \mid\equiv MS \overset{a^{xy} \bmod p}{\leftrightarrow} BS1 \tag{12}$$

From Eqs. (11 and 12), one can deduce that the proposed protocol achieves the goals of authentication and key distribution without bugs or redundancies. In the next section, comparison of the proposed protocol with other handover protocols is detailed.

5. Comparison of the Proposed Protocol with IEEE 802.16m and SCT-Based Protocols

The proposed protocol is compared with IEEE 802.16m and Fu et al. [12] protocols. The comparison will be undertaken according to: communication and computation overheads. In order to undertake the communication overhead comparison, the following parameters are defined: T_{EAP} is the delay for the full EAP authentication, T_w is the transmission latency between the MS and the BS, and T_a is the transmission latency between the BS and ASN-GW or latency between ASN-GW and AAA server. According to [4], the following values could be assumed: T_{EAP} = 1000 ms, T_w = 15 ms, and T_a = 20 ms. Table 1 shows a comparison of communication overheads concerning the proposed protocol, IEEE 802.16m and Fu et al. protocols. The table shows that the proposed protocol outperforms both the IEEE 802.16m and Fu et al. protocol for both inter-domain and intra-domain handover.

Table1:Comparison of communication overheads

IEEE 802.16mm	***Fu et al. Protocol***		***Proposed Protocol***	
	Intra-Domain Handover	***Inter-Domain Handover***	***Intra-Domain Handover***	***Inter-Domain Handover***
$T_{EAP}+T_a+3T_w$ = 1065 ms	$2T_a+3T_w$ = 85 ms	$T_{EAP}+T_a+3T_w$ =1065 ms	$2T_a+2T_w$ =70 ms	$3T_a+2T_w$ =90 ms

In order to carry out the computation overhead comparison, the following parameters are defined: T_{CMAC} is the time for CMAC operation, T_{hash} is the time to perform a hash operation, T_{DOT} is the time for a DOT operation, T_{sym} is the time to perform a symmetric operation, and T_{exp} is the time to calculate a modulo exponentiation. Table 2 shows a comparison of computation overheads concerning the proposed protocol, IEEE 802.16m and Fu et al. protocol. It has to be noted that the modulo exponentiation time is the most consuming: time among the abovementioned parameters. As stated in [12], the exponentiation time could be equal to 2 ms. Therefore, the computation overhead could be negligible compared to the communication overhead. As mentioned in the previous sections, the advantages of using Diffie-Hellman protocol are: to overcome the domino effect and also that both MS and BS shared a symmetric key which is only known to both of them which is not the case in both IEEE 802.16m and Fu et al. protocols. This leads to maintain the confidentiality of messages exchanged between both MS and BS. Although the computation overhead is higher than the other protocols, the latency performance (including communication and computation overheads) is still efficient. This is a result of the dramatically decrease in the communication overhead of the proposed protocol.

Table 2:Comparison of computation overheads

	MS	***BS***
IEEE 802.16mm	$2T_{DOT}$ + $3T_{CMAC}$	$2T_{DOT}$ + $3T_{CMAC}$
Fu. Et al Protocol	$4T_{DOT}$ + $3T_{CMAC}$	$4T_{DOT}$ + $3T_{CMAC}$+ $2T_{hash}$+ $2T_{sym}$
Proposed Protocol	$3T_{CMAC}$+$2T_{hash}$+ T_{exp}	$3T_{CMAC}$+$2T_{hash}$+ T_{exp}

6. Conclusions

Mobile WiMAX requires the re-authentication of the mobile stations as they change from one base station to another. IEEE 802.16m uses Extensible Authentication Protocol (EAP) for authentication and key management. This requires about 1000 ms, therefore, it could not support mobile WiMAX applications such as video conference. In literature, many solutions have been proposed to solve the re-authentication problem. These solutions solve only the problem of intra-domain handover (where the target base station and the serving base station are in the same domain). For inter-domain handover (where the target base station and the serving base station are in different domains), the MS must perform the full EAP authentication protocol which leads to an impractical latency for mobile WiMAX applications. In the present paper, a protocol that aims to overcome the inter-domain handover problem is proposed. The proposed protocol uses the Extensible Authentication Protocol (EAP) for authentication and key distribution. The proposed protocol is based on the use of hash functions and the Diffie-Hellman protocol to distribute the keys between mobile stations and base stations. In order to avoid the domino effect, the Diffie-Hellman components are distributed instead of the authentication key itself. The advantages of using Diffie-Hellman protocol are: to overcome the domino effect in case of BS compromise and also that both MS and BS shared a symmetric key which is only known to both of them which is not the case in both IEEE 802.16m and Fu et al. protocols. This leads to maintain the confidentiality of messages exchanged between both MS and BS. The proposed protocol is analyzed using the BAN logic to ensure that it achieves the goals of authentication and key distribution. The analysis shows that the proposed protocol achieves its goals without bugs or redundancies. Furthermore, the proposed protocol is compared with other handover protocols. It has been shown that, although the use of Diffie-Hellman protocol slightly increases the computation overhead, the proposed protocol still outperforms the other protocols. This is a result of the dramatically decrease in the communication overhead of the proposed protocol.

References

[1] IEEE 802.16e-2005, "Amendment for Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands", November 2005.
[2] IEEE 802.16-2004, "Air Interface for Broadband Wireless Access Systems", October 2004.
[3] IEEE 802.16m-2011, "Air Interface for Broadband Wireless Access Systems-Amendment 3: Advanced Air Interface", May 2011.
[4] J. Hur. H. Shim, P. Kim, H. Yoon and N. Song, "Security Considerations for Handover Schemes in Mobile WiMAX Networks", In WCNC Conference, 2008, pp. 2531-2536.
[5] IEEE 802.16m-2007, "Mobility Sensitive Master Key Derivation and Fast Re-authentication for 802.16m", February 2007.
[6] B. Aboba, "Fast Handoff Issues", IEEE-03-155r0-I, IEEE Working Group, 2003.
[7] WiMAX Forum, "Mobile WiMAX- Part I: A Technical Overview and Performance Evaluation", February 2006.
[8] H. M. Sun, Y. H. Lin, S. M. Chen, and Y. C. Shen, "Secure and Fast Handover Scheme Based on Pre-Authentication Method for 802.16/WiMAX Infrastructure Networks", In TENCON Conference, 2007, pp. 1-4.
[9] Y. Ohba, Q. Wu, and G. Zorn, "Extensible Authentication Protocol (EAP) Early Authentication Problem Statement", RFC 5836, April 2010. Available at: http://www.rfc-editor.org/rfc/rfc5836.txt.
[10] T. N. Nguyen, and M. Ma, "Enhanced EAP-Based Pre-Authentication for Fast and Secure Inter-ASN Handovers in Mobile WiMAX Networks", IEEE Transactions on Wireless Communications, Vol. 11, No. 6, June 2012, pp. 2173-2181.
[11] A. Fu, G. Zhang, Y. Zhang, and Z. Zhu, "GHAP: An Efficient Group-based Handover Authentication Mechanism for IEEE 802.16m Networks", Wireless Pers Communication, Published online: 05 August 2012.
[12] A. Fu, Y. Zhang, Z. Zhu, Q. Jing, and J. Feng, "An Efficient Handover Authentication Scheme with Privacy Preservation for IEEE 802.16m Network", Computers & Security, Vol. 31, No. 6, September 2012, pp. 741-749.
[13] Y. Kim, W. Ren, J. Jo, Y. Jiang, and J. Zhang, "SFRIC: a Secure Fast Roaming Scheme in Wireless LAN using ID-based Cryptography", In ICC Conference, 2007, pp. 1570-1575.
[14] C. Zhang, R. Liu, P. Ho, and A. Chen, "A Location Privacy Preserving Authentication Scheme in Vehicular Network", In WCNC Conference, 2008, pp. 2543-2548.
[15] J. Cao, H. Li, M. Ma, Y. Zhang, and C. Lai, "A Simple and Robust Handover Authentication Between HeNB and eNB in LTE Networks", Computer Networks Journal, Vol. 56, No. 8, May 2012, pp. 2119-2131.
[16] M. Nakhjiri, "Use of EAP-AKA, IETF HOKEY and AAA Mechanisms to Provide Access and Handover Security and 3G-802.16m Internetworking", In PIMRC Conference, 2007, pp. 1-5.
[17] V. Narayanan, and L. Doneti, "EAP Extensions for EAP Re-authentication Protocol (ERP)", IETF RFC 5296, August 2008.
[18] Y. Ohba, S. Das, and A. Dutta, "Kerberized Handover Keying: a Media Independent Handover Key Management Architecture", In MobiArch Conference, 2007, pp. 1-7.
[19] M. Rafa, P. Fernando, O. Yoshihiro, B. Fernando, and F. Antonio, "A Kerberized Architecture for Fast Re-authentication in Heterogeneous Wireless Networks", Mobile Network Applications Journal, Vol. 15, No. 3, 2010, pp. 392-412.
[20] C. Politis, K. Chew, N. Akhtar, M. Georgiades, R. Tafazolli, and T. Dagiuklas, "Hybrid Multilayer Mobility Management with AAA Context Transfer Capabilities for all IP Networks", IEEE Wireless Communications Journal, Vol. 11, No. 4, 2004, pp. 76-88.
[21] C. Huang, and J. Li, "A Cluster Chain Based Context Transfer Mechanism for Fast Basic Service Set Transition in the Centralized Wireless LAN Architecture", Wireless Communication Mobile Computing Journal, Vol. 9, No. 10, 2009, pp. 1387-1401.
[22] J. Loughney, M. Nakhjiri, C. Perkins, and R. Koodli, "Context Transfer Protocol (CXTP)", IETF RFC 4076, July 2005.
[23] R. Housley, and B. Aboba, "Guidance for Authentication, Authorization and Accounting (AAA) Key Management", IETF RFC 4962, July 2007.
[24] IEEE 802.16m-09/0034r3, "IEEE 802.16m System Description Document", June 2010.
[25] M. Burrows, M. Abadi, and R. Needham, "A Logic of Authentication", ACM Transaction on Computer Systems, Vol. 8, No. 1, February 1990, pp.18-36 .

Heba Kamal Aslan is an Associate Professor at Electronics Research Institute, Cairo-Egypt. She received her B. Sc. degree, M. Sc. degree and Ph. D. degree in Electronics and Communications Engineering from the Faculty of Engineering, Cairo University, Egypt in 1990, 1994 and 1998 respectively. Aslan has supervised several masters and Ph. D. students in the field of computer network security. Her research interests include: Key Distribution Protocols, Logical Analysis of Protocols and Wireless Security.

A Cipher based on Multiple Circular Arrays

Pushpa R. Suri[1], Sukhvinder Singh Deora[2]

**[1]Department of Computer Sc. & Applications,
Kurukshetra University, Kurukshetra,
Haryana, India-136119.**

**[2]Dept. of Computer Applications,
N. C. Institute of Computer Sciences, Israna,
Panipat, India-132107.**

Abstract

Security of information communicated over the internet has been a challenged area. Data is encrypted by using some cryptographic algorithms using symmetric/asymmetric/public key algorithms for its protection. During literature survey of such algorithms, it was observed that there has been use of one or two dimensional array structures in development of cryptographic algorithms. This paper explains the use of alternative structure made of multiple circular arrays in development of a block cipher based on symmetric variable length key encryption for diffusion of information to a larger extent. It uses operations like circular shift operations, merge-swap and XOR to encrypt the data. The key used for encryption algorithm is generated in parts using Random Number Generator (RNG) operated with XOR operations with its previous sub-key bits at both ends. It may also be transferred along with the ciphertext to intended receiver. The cryptographic algorithm generates scrambled data with above mentioned operations for confusion-diffusion of bits to a larger extent of plaintext data using easy to compute operations for efficient security. The algorithm may also work in asymmetric mode using same direction of rotations. The research work presented in the paper can be used for data security over networks.

Keywords: *Arrays, composite data structures, Cryptographic algorithms, data security, encryption-decryption, symmetric key cryptography, Random number generation, Random Number Generators (RNGs).*

1. Introduction

Security of information on computers & networks has been a challenged area. Cryptography is the branch of computer science that deals with hiding information for secure communication of data. It uses encryption-decryption algorithms (Cryptographic algorithms) to convert readable information to some non-readable form (ciphertext). It uses many mathematical operations along with data structures to encrypt and decrypt the information [1,2]. Sender of information shares decryption technique that is required to recover the original information (plaintext). Literature survey on cryptography and security, it was observed that researchers have proposed algorithms using one dimensional or two dimensional arrays only. This paper presents use of circular shift, merge-swap and XOR operations on Multiple Circular Arrays (MCAs) as a symmetric encryption key algorithm [3]. The key used is generated in steps by an agreed upon Random Number Generator at both ends or is communicated along with the ciphertext to its intended recipient as required. The recipient uses the inverse operations in exactly the opposite order to receive back the plaintext from the ciphertext.

2. Cryptographic Algorithms

Cryptographic algorithms are the basic building blocks in cryptography. They are used in various security applications and protocols. A cryptographic algorithm or cipher is a mathematical function used to encrypt information [4]. Generally, there is an alternate function, unique against the encryption function that is used for decryption of the information. If the way that the algorithm works is kept secret, the algorithm is called restricted algorithm. The quality control and standardization for such an algorithm is generally not so good. It is required that the security of algorithm does not depend on the secrecy of its working. In fact the use of operations in the algorithms should be such that the probability of decrypting the information without actually having the decryption key should be almost negligible or very difficult.

The algorithms are mainly divided into three categories on the basis of the fact of using the same or different key for encryption and decryption:

1. Symmetric Algorithms
2. Asymmetric Algorithms
3. Public-Key Algorithms

2.1 Symmetric Algorithms

In these algorithms, the sender and receiver agree on a key before they communicate. Cryptographic algorithms that use same key, K for encryption and decryption of the information are called symmetric algorithm(s) and such technique of communicating securely is called Symmetric Key Cryptography (SKC).
Encryption and Decryption are denoted as Eq. (1) & (2)

$$E_K(M) = C \quad (1)$$

$$D_K(C) = M \quad (2)$$

The strength of a symmetric key cryptography depends on the strength of the algorithm and the length of key used, assuming that the algorithm is perfect, i.e. there is no other way to break the system than the brute-force attack. Symmetric-key systems like Word Auto Key Encryption (WAKE), SEAL, IDEA, RC4, RC5 etc are simpler and faster [1,5,6,7,8,9,10]. Most recently, many symmetric ciphers based on matrix and operations on it were proposed [11].

However, the quest for generating faster symmetric algorithms based on faster hardware properties is also going on [12]. Researchers all over the world are reviving the way symmetric ciphers have been in use in the past [13,14,15]. On the basis of operations performed on single bit or a group of bits, the symmetric ciphers may be divided into stream algorithms or block algorithms respectively.

2.2 Asymmetric Key Algorithms

The algorithm is designed such that there is a separate key for sender and receiver. The key K is used for encryption and K' is used while decryption of the information. Such an encryption is called Asymmetric Key Cryptography (AKC). In this case, there is a combination of keys (K, K') used to encrypt using encryption algorithm (E_K) and decrypt using decryption algorithm ($D_{K'}$). A prior knowledge to be shared in this case may be the rule that can derive K' if K is known or vice-versa or combination keys and key K used while encryption. Encryption and Decryption are denoted as Eq. (3) & (4):

$$E_K(M) = C \quad (3)$$

$$D_{K'}(C) = M \quad (4)$$

2.3 Public Key Algorithms

These are the asymmetric algorithms that are designed in such a way that the key used to encrypt is different than that used for decryption. The encryption key (K) is available to all so that everyone (known as Public Key) can encrypt the information using encryption algorithm (E_K). But only specific person possesses the decryption key (K'), known as Private Key, that is used to decrypt the message using decryption algorithm ($D_{K'}$). It is not easy to predict the decryption key given the knowledge of encryption key and the algorithm. It is not easy to predict the decryption key given the knowledge of encryption key.
Encryption and Decryption are denoted as Eq. (5) & (6):

$$E_K(M) = C \quad (5)$$

$$D_{K'}(C) = M \quad (6)$$

Diffie and Hellman introduced this encryption scheme in 1976 where each person gets a pair of keys, called the public key and the private key. Public key of persons is published while keeping the private key(s) as secret. Public key encryption avoids the problem of securely exchanging the keys because the public key can be distributed in any manner, and the private key is never transmitted. Messages are encrypted using intended recipient's public key which can only be decrypted using his private key. The need for sender and receiver to share secret information (keys) via some secure channel is eliminated. It may be used for authentication, confidentiality, integrity and non-repudiation. RSA encryption algorithm is a public-key cryptosystem [16]. Researchers have been developing many public-key cryptographic algorithms but they are slow due to complex operations carried out based on hard to solve mathematical problems.

The next sections will detail about the proposed cipher that uses circular arrays as mathematical structure and new operation of merge-swap to produce a ciphertext. It also elaborates the structure, operations to be performed, key and its length, detailed encryption-decryption algorithms, the result and analysis of its sample runs, its possible variants. It is explained using a symmetric key and is a block cipher.

3. The Proposed Cipher

This section explains the use of Multiple Circular Arrays (MCAs), a composite structure that each circular array has double the size of its previous circular array. The number of such circular arrays to be used in the structure depends upon the size of data on which the encryption-decryption is to be applied.

3.1 The Structure

The Multiple Circular Arrays (MCAs) structure has been thought of as multiple Rotor Plates numbered from 1 to n from innermost to outside the structure. A example MCA is shown in Fig. 1 that contains three circular arrays and size of innermost array is 4. The number of elements in the innermost circular array is shown with 4 elements with data items filled from 1 to 4. The subsequent circular arrays have size of 8 and 16 having 5 to 12 in second and 13 to 28 in third circular array and so on. There are total of 3 circular arrays having a total of 4(1+2+4)=28 elements (refer to Fig. 2).Persons under communication may decide the size of innermost array and number of such arrays in the MCA structure before actual communication.

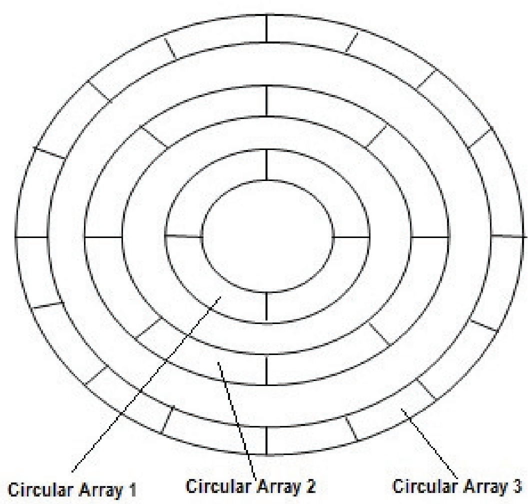


Fig. 1. View of Multiple circular Array

3.2 Operations

The size of the structure will enable some more operations, apart from the traditional operations like XOR and circular right/left shift operations on its elements. The shape of circular arrays is like a rotar plate and hence circular right/left shift operation is shifting the elements of the circular array by some number of positions. E.g. Fig. 3 shows the resultant of the MCAs after the second circular array is rotated right by two positions.

Swapping of variables is operation often used in programming. It is known that swapping can also be viewed as XOR operation. That is, when we XOR any two items, say A and B so that its resultant value is C, refer to Eq. (7). Now, if we XOR C with either of the two initial values, the other will be produced as a resultant value, refer to Eq. (8).

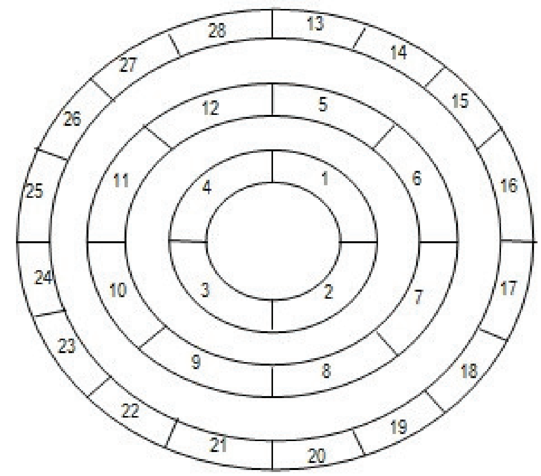


Fig. 2. Data sequence shown

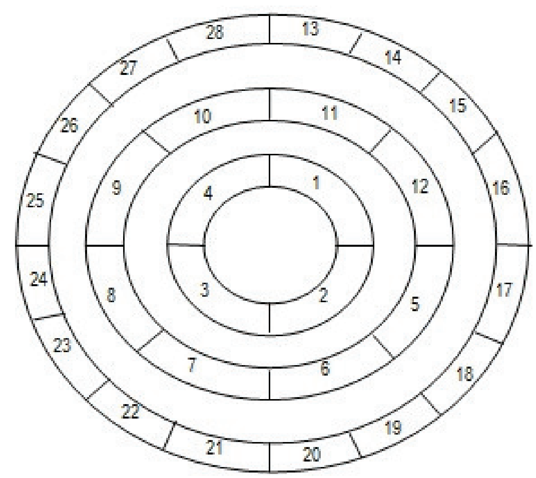


Fig. 3. Multiple Circular Arrays after Circular Rotate operation on 2nd Plate by 2 positions

A XOR B = C and (7)
A XOR C = B or B XOR C = A. (8)

Similarly, swapping can be used as an involution function, i.e. $f^{-1}=f$. Thus, if we have two variables A & B such that A=2 and B=3. Applying swapping operation

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 1, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

once will result into A=3 and B=2. If swapping is applied again on the two variables A & B, it will lead to A=2 and B=3. We are proposing a merge-swap operation in the proposed cipher that uses swapping of elements of a particular circular array with the next circular array starting. Swapping may start at a particular index value and can be done with/without gap.

Merge-swap operation of 1st circular array with 2nd circular array of Fig. 2 has been explained without gap in Fig. 4 and with gap option in Fig. 5.

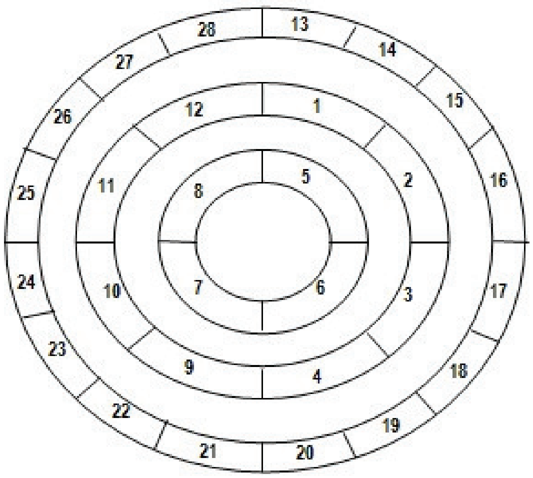


Fig. 4. Sample MCAs after Merge-swap of 1st Circular Array with 2nd Circular Array starting at position 1 in 2nd Circular Array without gap

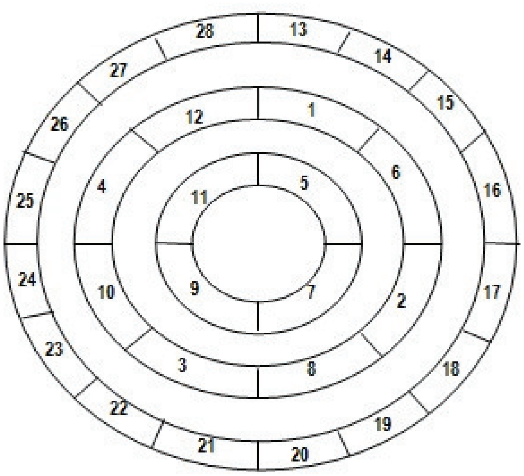


Fig. 5. Sample MCAs after Merge-swap all the elements of 1st Circular Array with 2nd Circular Array starting at position 1 in 2nd Circular Array with gap

3.3 Key, its generations & use

The cipher has been designed in such a manner that one can take a variable sized plaintext by deciding the total number of circular arrays to be used for storing the plaintext initially by padding 0s in the end, if required. Thus, the variable key length is possible for the suggested cipher depending upon the number of circular arrays used in the structure. The general approach to generate the key bits, number of bits used and bit positions to be used for specific operations is discussed in this section.

The proposed key is generated step-wise for each circular array. If we are to encrypt plaintext contained in a multiple circular arrays with innermost array of size = 8 bits, then the first sub-key G_1 of length 8 bits is generated using a Random Number Generator (RNG). It is used as S_1, sub-key for encryption of innermost circular array. For the next circular array, G_2, a set of 8 bit random number is again generated. This is XORed with the previous sub-key of length 8 bits and appended to the resultant 8 bits thereby making it now 16 bits sub-key for 2nd rotar plate. Table 1 provides the the details of generating the sub-keys for each rotar plate.

Table 1: Key Generation w.r.t. Circular Array Number

Circular Array Number	Length of sub-key generated (G_i)	Sub-key used for encryption S_i (i = plate number)
1	G_1=8 bits	$S_1=G_1$
2	G_2=8 bits	S_2= (S_1 XOR G_2)appendG_2
3	G_3=16 bits	S_3= (S_2 XOR G_3)appendG_3
4	G_4=32 bits	S_4= (S_3 XOR G_4)appendG_4
and so on for next rotor plates/circular arrays …		
i	G_i=(i-1)*8	S_i= (S_{i-1} XOR G_i)append G_i

The cipher uses some number of bits for applying operations discussed in section 3.2. The total number of bits to be used for particular Circular Array Number is provided in the Table 2.

Table 2: Number of bits to be used during Encryption/Decryption w.r.t. Circular Array Number

Circular Array Number	Total No. of bits					
	Newly Generated	Used for				
		Merge-swap		Rotate operation	Rotate and Merge-swap	XOR
		Gap	Start Position			
1	8	1	4	3	8	8
2	8	1	5	4	10	16
3	16	1	6	5	12	32

4	32	1	7	6	14	64
and so on for next rotor plates/circular arrays…						
i	2^{i+1}	1	3+i	3+(i-1)	7+2i-1	2^{i+2}

A 8 bit sub-key is used for applying operations on innermost circular array bits in some order; say circular rotate, Merge-swap and XOR key bits with the circular array bits. The bit numbers that were used for applying the three operations in the test runs of the proposed cipher is given in Table 3.

Table 3: Key bit positions used for various operations during Encryption/Decryption w.r.t. Circular Array Number

Circular Array Number	Length of sub-key used	Rotate bits	Merge-swap bits	
			With Gap bit	Position of bit where to start swap in next Plate
1	8	5 to 7	0	1 to 4
2	16	12 to 15	6	7 to 11
3	32	27 to 31	20	21 to 26
4	64	58 to 63	50	52 to 57
and so on for next rotar plates…				
i	2^{i+2}	$(2^{i+2}$-(i+2)) to $(2^{i+2}$-1)	$(2^{i+2}$-2(i+3))	$(2^{i+2}$-2(i+2)-1) to $(2^{i+2}$- (i+2)-1)

Sub-keys may be repeatedly used for the same circular array i, after circular right shift operation on the sub-key S_i by 3 positions. The number of repetitions used for applying operations on circular array bits will increase the overall security of the algorithm. While decryption, the sub-key S_i is rotated circular left shift operation by the same number of bit positions.

3.2 Encryption Algorithm

The proposed cipher has been designed using Multiple Circular Arrays (MCAs) with number of such arrays equal to n. The plaintext may be inserted into the MCAs starting from innermost to outermost circular array taken in order of the index values. The operations carried out on the circular array "i" under consideration may be repeated for some odd number of times as per the requirements of level of data security to be achieved. The algorithm uses number of bits and bit positions of key as discussed in table 1, 2 and 3 of section 3.3. Variable i denotes the number, CA_i denotes i-th Circular Array, Count variable is used to calculate the number of iterations performed on CA_i, $\oplus$ denotes append/concatenate operation, << and >> denotes the circular left shift and circular right shift operations respectively in the Algorithms given below:

Encryption Algorithm

1. Set Count :=0.
2. Repeat Steps 3 to 10 For i=1 to n, do the following for circular array CAi:
3. Generate the sub-key G_i for i-th circular array.
4. If (i <> 1) then
   Update sub-key $S_i = (S_{i-1}$ XOR $G_i) \oplus G_i$
   Else
   Set $S_i = G_i$
   [End of If condition, step no. 4]
5. Obtain the circular right bit values as discussed in table 1, 2 & 3.
6. Perform circular right shift operation on CA_i as per the value of bit positions discussed in Section 3.3
7. If (i <> n) then
   Obtain the merge-swap bit values
   Perform merge-swap operation between CA_i and CA_{i+1}
   [End of If condition, step no. 7]
8. Perform CA_i XOR S_i.
9. Count:=Count + 1.
10. If (Count < 3) then
    $S_i = (S_i << 3)$.
    Else
    Set Count := 0
    [End of If condition, step no. 10]
    [End of For loop, step no. 2]
11. End

At the end of the encryption algorithm, the CA_is will contain the ciphertext that can be collected in the order that they were inserted into the CA_is. The ciphertext so obtained along with the key S_i used in last circular array iteration needs to be communicated to its recipient.

The recipient may start the decryption using the decryption algorithm that uses the key S_i for i-th circular array CA_is. The detailed decryption algorithm is as follows:

Decryption Algorithm

1. Set Count :=0. Set i=n.
2. Repeat Steps 3 to 10 While i >= 1, do the following for circular array CAi:
3. If (Count = 0) then
   Use the sub-key S_i for i-th circular array.
   Else
   Set $S_i = S_i >> 3$.
   [End of If condition, step no. 3]

4. Perform CA_i XOR S_i.
5. If (i <> n) then
   Obtain the merge-swap bit values
   Perform merge-swap operation between CA_i and CA_{i+1}
   [End of If condition, step no. 7]
6. Obtain the circular right bit values as discussed in table 1, 2 & 3.
7. Perform circular left shift operation on CA_i as per the value of bit positions discussed in Section 3.3
8. Set Count = Count + 1.
9. If (Count = 2 and i > 1) then
   Set G_i = right half key bits of S_i and
   Set S_{i-1} = left half key bits of S_i.
   Obtain sub-key S_{i-1}= (S_{i-1} XOR G_i)
   Set Count = 0.
   [End of If condition, step no. 8]
10. Set i := i – 1.
    [End of While loop, step no. 2]
11. End

4. Some Test Results

The proposed algorithm was implemented in C language. Multiple Circular Arrays of were initialized with half 0s and half 1s as per the size of the arrays. The operations as discussed in Section 3 were applied for certain number of rounds and the resultant bits were collected by reading the arrays from innermost to outermost. Randomly selected subsequence of bits were chosen and tested for the amount of randomness produced after the shuffling of plaintext bits as per the NIST specification tests [17]. The tests were applied to know how much shuffling of bits has been done by the operations. The performance of the above discussed algorithm in some selected tests was recorded and is explained in the following subsections:

4.1 Monobit Test

This test is used to determine the number of 0s and 1s in a randomly selected bit sequence of ciphertext is same or not. If the resultant sequence is a random sequence, then any arbitrarily selected bits must have equal proportion of 0s and 1s. The objective of the test is to assess closeness of the fraction of ones to ½. That is, thenumber of zeros and ones in a sequence should be almost equal. It uses parameters n - the length of the bit string, $\varepsilon = \varepsilon_1, \varepsilon_2, \varepsilon_3,\dots, \varepsilon_n$ the sequence of bits. It then calculates S_{obs} – the absolute value of the sum of X_i (where $X_i = 2\varepsilon - 1 = \pm 1$) given by Eq. (9).

$$S_{obs} = \frac{|S_n|}{\sqrt{n}} \qquad (9)$$

The p-value is then computed as given by Eq. (10),

$$p-value = erfc\left(\frac{S_{obs}}{\sqrt{2}}\right) \qquad (10)$$

where *erfc* is the complementary error function.

The tests were applied on sample data of 7 and 8 circular arrays. The p-value calculated after 5, 6 or 7 rounds were evaluated. The details of the min., max., average p-value and number of tests failed and passed is provided(see Table 4 & 5). Tests were applied on structures with some number of circular arrays after applying certain number of rounds/iterations (see Table 4 & 5).

Table 4: Test Results of Monobit Test on 7 Circular Arrays

S No	NoOfRounds, NoOfCAs		
	5, 7	6, 7	7, 7
	p-value	p-value	p-value
1	0.657969	0.282297	0.087705
2	0.230139	1	0.109599
3	1	0.423711	0.689157
4	0.841481	0.548506	0.689157
5	0.027807	0.423711	0.689157
6	0.230139	0.027807	0.317311
7	0.548506	0.689157	0.027807
8	0.230139	0.109599	0.423711
9	0.689157	0.689157	0.423711
10	0.841481	0.423711	0.841481
11	0.841481	0.548506	0.423711
12	0.949571	0.282297	0.113846
13	0.423711	0.548506	0.161513
14	0.423711	0.841481	1
15	1	0.423711	0.689157
16	0.009322	0.109599	0.423711
17	0.548506	0.230139	0.317311
18	0.230139	0.689157	0.009322
19	0.230139	0.109599	0.317311
20	0.423711	0.841481	0.423711
21	1	0.841481	0.689157
22	0.548506	0.841481	0.841481
Min	0.009322	0.027807	0.000689
Max	1	1	1
Average	0.542073	0.496595	0.446975
Result:	Count	Count	Count
FAIL	1	0	1
PASS	21	22	21

The results of the test indicate that more than 95% of times the resultant ciphertext contained equal proportion of 0s and 1s again. It indicates that operations have shuffled, XORed and merge-swapped the bits, still the basic nature of bits have not been lost. Thus there is a good shuffling of the elements of the plaintext.

Table 5: Test Results of Monobit Test on 8 Circular Arrays

S No	NoOfRounds, NoOfCAs		
	5, 8	6, 8	7, 8
	p-value	p-value	p-value
1	0.375921	0.949571	0.282297
2	0.548506	0.423711	0.689157
3	0.689157	1	0.230139
4	0.841481	0.689157	0.548506
5	0.423711	0.689157	0.841481
6	0.841481	0.689157	0.689157
7	0.109599	0.230139	0.548506
8	0.071861	1	0.548506
9	0.689157	0.161513	0.423711
10	0.0455	0.689157	0.0027
11	0.423711	1	0.109599
12	0.569214	0.612882	0.447884
13	0.548506	0.548506	0.689157
14	0.071861	0.689157	0.317311
15	0.841481	0.841481	0.109599
16	0.841481	0.230139	0.841481
17	0.548506	0.071861	0.548506
18	0.841481	0.689157	0.317311
19	0.548506	0.016395	0.230139
20	0.548506	0.841481	0.841481
21	0.689157	0.423711	0.423711
22	0.841481	0.109599	0.317311
Min	0.0455	0.016395	0.000689
Max	0.841481	1	1
Average	0.543194	0.572542	0.452748
Result:	Count	Count	Count
FAIL	0	0	1
PASS	22	22	21

4.2 Frequency within a Block Test

The focus of this test is on the proportion of 1s within M-bit blocks of the data. It evaluates that the number of 1s in a M-bit data block is approximately M/2. This test uses M-the length of each block, n – the length of the bit string and sequence of bits $\varepsilon = \varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots, \varepsilon_n$. Firstly, non-overlapping blocks N are formed given by Eq. (11):

$$N = \left\lfloor \frac{n}{M} \right\rfloor \qquad (11)$$

Now, use Eq. (12) to determine the proportion π_i of ones in each M-bit block.

$$\pi_i = \frac{\sum_{j=1}^{M} \varepsilon_{(i-1)} M + j}{M} \qquad (12)$$

and compute the χ^2 statistic as per Eq. (13) to further obtain the p-value using Eq. (14).

$$\chi^2(\text{obs}) = 4M \sum_{i=1}^{N} \left(\pi - \frac{1}{2}\right)^2 \qquad (13)$$

$$\text{p-value} = igamc\left(\frac{N}{2}, \chi^2(obs)/2\right) \qquad (14)$$

Table 6: Test Results of Frequency within a Block Test on 7 Circular Arrays

S No	NoOfRounds, NoOfCAs		
	5, 7	6, 7	7, 7
	p-value	p-value	p-value
1	0.999438	0.999887	0.99982
2	0.991468	0.999993	0.999988
3	0.964295	0.998474	0.998821
4	0.911413	0.999951	0.999856
5	0.834308	0.998821	0.99982
6	0.739918	0.999777	0.997823
7	0.637119	0.999107	0.999934
8	0.534146	0.999982	0.999777
9	0.437274	0.999438	0.999988
10	0.350485	0.99923	0.999887
11	0.999524	0.998474	0.99472
12	0.999107	0.99934	0.998058
13	0.998971	0.999438	0.999951
14	0.998275	0.998058	0.998821
15	0.999668	0.998971	0.999438
16	0.999601	0.999668	0.998655
17	0.999107	0.997568	0.999668
18	0.999668	0.999107	0.999974
19	0.999934	0.99934	0.998275
20	0.995969	0.999726	0.999524
Min	0.350485	0.997568	0.99472
Max	0.999934	0.999993	0.999988
Average	0.851823	0.999178	0.998978
Result:	Count	Count	Count
FAIL	0	0	0
PASS	20	20	20

Table 7: Test Results of Frequency within a Block Test on 8 Circular Arrays

S No	NoOfRounds, NoOfCAs		
	5, 8	6, 8	7, 8
	p-value	p-value	p-value
1	0.999726	0.998821	0.999913
2	0.99934	0.999668	0.999524
3	0.999934	0.999913	0.996677
4	0.999913	0.999777	0.997823
5	0.999438	0.999988	0.995578
6	0.999668	0.999777	0.99923
7	0.998655	0.996996	0.999934
8	0.998474	0.999601	0.99934
9	0.996335	0.99934	0.997823
10	0.998058	0.985339	0.999668
11	0.999934	0.999856	0.999668
12	0.998655	0.999887	0.998971
13	0.999601	0.999726	0.999668
14	0.997568	0.999524	0.999726
15	0.999777	0.985339	0.999438
16	0.999668	0.999964	0.999934
17	0.999974	0.999524	0.999601
18	0.99982	0.999601	0.999913
19	0.998474	0.99472	0.999934
20	0.999993	0.998655	0.999951
Min	0.996335	0.985339	0.995578
Max	0.999993	0.999988	0.999951
Average	0.99906	0.997334	0.998993
Result:	Count	Count	Count
FAIL	0	0	0
PASS	20	20	20

The parameters of the frequency within a block test are evaluated using (14). The resulted values show that all tests were passed by the proposed cipher for 7 and 8 Circular Array block (see Table 6 & 7). It confirms that the operations applied on the structure are scrambling the data bits amongst themselves.

4.3 Runs Test

The focus of this test is the total number of runs of 1s in the sequence. A run is an uninterrupted sequence of identical bits. A run of length k consists of exactly k identical bits that are bounded before and after with a bit of the opposite value. The main purpose of the runs test is to determine whether the number of runs of one and zeros of various lengths is as expected for a random sequence. Moreover, it checks the oscillations of zeros and ones are more or low in number. For a n - length of bit string, let ε - denotes the sequence of bits $\varepsilon = \varepsilon_1\varepsilon_2\varepsilon_3 \ldots \varepsilon_n$. The parameter $V_n(obs)$ – denotes the total number of runs (zero runs + one runs) across all n bits evaluated given by Eq. (16) and

$$\tau = \frac{2}{\sqrt{n}} \tag{15}$$

$$V_n(\text{obs}) = \sum_{k=1}^{n-1} r(k) + 1 \tag{16}$$

where r(k)=0 if $\varepsilon_k = \varepsilon_{k+1}$ r(k)=1 otherwise.

The p-values is then calculated given by Eq. (17)

$$p - value = erfc\left(\frac{\left| V_n(obs) - 2n\pi(1-\pi) \right|}{2\sqrt{2n\pi}(1-\pi)} \right) \tag{17}$$

Table 8: Test Results of Runs Test on 7 Circular Arrays

S No	NoOfRounds, NoOfCAs		
	5, 7	6, 7	7, 7
	p-value	p-value	p-value
1	0.79047	0.041017	0.996745
2	0.080606	0.421211	0.548506
3	0.161513	0.284236	0.537243
4	0.894201	0.500798	0.284236
5	0.443224	0.505677	0.919542
6	0.385117	0.308855	0.894201
7	0.505677	0.379678	0.919542
8	0.948642	0.545683	0.640508
9	0.689157	0.844549	0.308855
10	0.100609	0.425847	0.83829
11	0.795064	0.161513	0.576645
12	0.071861	0.203323	0.828718
13	0.23151	0.812771	0.676922
14	0.4518	0.948642	0.828718
15	0.643606	0.739835	0.266521
16	0.571368	0.223244	0.739835
17	0.643606	0.278479	0.948642
18	0.853782	0.308855	0.948642
19	0.545683	0.346178	0.421211
20	0.315185	0.52328	0.503957
Min	0.071861	0.041017	0.266521
Max	0.948642	0.948642	0.996745
Average	0.506134	0.440183	0.681374
Result:	Count	Count	Count
FAIL	0	0	0
PASS	20	20	20

Table 9: Test Results of Runs Test on 8 Circular Arrays

S No	NoOfRounds, NoOfCAs		
	5, 8	6, 8	7, 8
	p-value	p-value	p-value
1	0.29846	0.503957	0.308855
2	0.853782	0.841481	0.338168
3	0.545683	0.413753	0.812771
4	0.79047	0.676922	0.551016
5	0.108573	0.413753	0.537243
6	0.13511	0.338168	0.661694
7	0.937395	0.230139	0.661694
8	0.432303	0.154473	0.948642
9	0.531971	0.853782	0.322658
10	0.140635	0.0455	0.379678
11	0.479523	0.384538	0.987214
12	0.166685	0.551016	0.52328
13	0.160152	0.443224	0.197399
14	0.423711	0.028706	0.23151
15	0.853782	0.338168	0.869265
16	1	0.761867	0.812771
17	0.432303	0.212076	0.835168
18	0.52328	0.171022	0.223244
19	0.00511	0.167799	0.714876
20	0.100609	0.588131	0.09399
Min	0.00511	0.028706	0.09399
Max	1	0.853782	0.987214
Average	0.445977	0.405924	0.550557
Result:	Count	Count	Count
FAIL	1	0	0
PASS	19	20	20

The test is applicable when all the frequency tests are cleared by some sequence of bits. Since the proposed cipher has cleared all the tests in Section 4.2, we may apply this test. This test was applied on 7 and 8 circular array structures with different number of rounds. The evaluated p-values show that most of the tests were passed (see Table 8 & 9).

4.4 Random Excursion Test

This test focuses on the number of cycles having exactly K visits in a cumulative sum random walk. The cumulative sum random walk is derived from partial sums after the (0,1) sequence is transferred to the appropriate (-1, +1) sequence. A cycle of a random walk consists of a sequence of steps of unit length taken at random that begin at and return to the origin. The objective of the test is to find out if the number of visits to a particular state within a cycle deviates from an random sequence. This test is a series of eight tests, one test and conclusion for each of the states: -4, -3, -2, -1 and 1, 2, 3, and 4.

Let n denote the length of bit string, ε - denotes the sequence of bits $\varepsilon = \varepsilon_1 \varepsilon_2 \varepsilon_3 \ldots \varepsilon_n$. The 0s and 1s of the input sequence (ε) are changed to values -1 and +1 using $X_i=2\varepsilon_i-1$. Then partial sums S_i are computed given by Eq. (18)

$$\begin{aligned} S_1 &= X_1 \\ S_2 &= X_1 + X_2 \\ S_3 &= X_1 + X_2 + X_3 \\ &\vdots \\ S_n &= X_1 + X_2 + X_3 \ldots X_n \end{aligned} \tag{18}$$

A new sequence S' is formed by attaching zeros before and after the set S. S'=0, S_1, S_2, S_3,… S_n. Total number of zero crossings in S' are calculated represented as J (the number of cycles in S'). Using the calculated values, p-value is computed and checked. The results (see Table 10 & 11) show that most of the tests have passed.

Table 10: Test Results of Random Excursion Test on 7 Circular Arrays

Test Result	NoOfRounds, NoOfCAs		
	5 7	6 7	7 7
PASS/FAIL	Count	Count	Count
FAIL	1	0	0
PASS	359	360	360
Min	0.000407	0.157299	0.157299
Max	1	1	1
Average	0.680464	0.644273	0.665769

Table 11: Test Results of Random Excursion Test on 8 Circular Arrays

Test Result	NoOfRounds, NoOfCAs		
	5 8	6 8	7 8
PASS/FAIL	Count	Count	Count
FAIL	0	0	0
PASS	360	360	360
Min	0.220672	0.157299	0.220672
Max	1	1	1
Average	0.658914	0.665038	0.652101

5. Analysis

Our proposed symmetric key algorithm based on multiple circular arrays use the operations of circular rotate, merge-swap and XOR three times on each of the circular array starting from the innermost array. Circular rotate operation shifts the plaintext data bits of CA_i by some integer value thereby ensuring change in the actual value. Merge-swap operation shuffles the plaintext data bits of some circular array, CA_i, with some subset of next circular array, CA_{i+1}. Lastly, the XOR operation with the sub-key S_i will change some of the data bits according to bit value of S_i. Repeating the process for some number of times ensures that scrambling of plaintext data bits to ciphertext. It is proposed that the repetition of the process must be done odd number of times so that the original plaintext is not produced as the ciphertext due to even iteration working as an inverse to the previous odd iteration. Sub-key is rotated circularly by 3 bits (or for any odd number) before applying further iterations on CA_i, thus reducing the probability that second iteration on CA_i becomes exactly inverse of the first iteration is very low.

6. Possible Variants and Future Scope

The proposed cipher is a variable length block cipher has a property to select as many multiple arrays as per the size of data (appending some dummy information bits/bytes) and may use variable length of key size (depending upon size of MCAs). It may be used with some secret S-box used for substituting step on data values to further increase the complexity of the encryption/decryption. This step will convert the cipher into a complex Feistel cipher. The cipher may be used in composite ciphers in which more than one type of cryptographic algorithms may be applied in sequence to produce harder algorithms. The operations suggested in the paper may be used in authentication and verifiability of data in data files using some modifications also.

7. Conclusions

The security of the cipher is dependent upon the symmetric key, number of iterations used for each circular array and rotation policies. Thus the attacker has to predict all three parameters well so as to decipher the ciphertext. The chance of making a brute force attack is very complex as there is a possibility of arriving at many number of possible plaintext data bits in that case. Error in prediction of a single bit value of the sub-key may affect many values in the ciphertext. This error will be further propagated to more plaintext bits due to merge-swap operation with their immediate successor circular arrays, if applicable. To sum up, the proposed symmetric block cipher has a good security with variable key length option. The key length to be used and number of iterations may also be increased/decreased as per agreed upon requirements of the entities involved in communication. Thus total number of possible attempts to be checked for n rounds will be 2^{n+2}. The complexity of decrypting the ciphertext to original plaintext increases with the increase in value of n. Moreover, error in prediction of one bit value will lead to multiple errors propagated to next rounds of iterations, a selective bit attacks may also not prove good. In nutshell, the proposed cipher is a flexible symmetric key cipher with variable key length option.

References

[1] Bruce Schneier, Applied Cryptography: John Wiley & Sons (Asia) Pte Ltd, ISBN 9971-51-348-X, 2006.

[2] William Stallings, Cryptography and Network Security, Principles and Practices, Fourth Edition: Pearson Education, 2006.

[3] Seymour Lipschutz, Data Structures: Tata McGraw Hill Publishing Company Ltd., 2005.

[4] Jonathan Katz and Yehuda Lindell, Introduction to Modern Cryptography, CRC Press, New York, pp. 3-4, 2007.

[5] D.J. Wheeler, "A Build Data Encryption Algorithm", Fast Software Encryption, Cambridge Scurity Workshop Proceedings, Springer-Verlag, 1994, pp. 127-134.

[6] P. Rogaway and D. Coppersmith, "A Software-Oriented Encryption Algorithm", Fast Software Encryption, Cambridge Security Workshop Proceedings, Springer-Verlag, 1994, pp. 56-63.

[7] X. Lai, "Detailed Description and a Software Implementation of the IPES Cipher", unpublished manuscript, 8 Nov. 1991.

[8] X. Lai, "On the Design and Security of Block Ciphers", ETH Series in Information Processing, Vol. 1, Konstanz: Hartung-Gorre Verlag, ETH No. 9752, Doctor of Technical Sciences, Xidian University, Xian, China, 1992.

[9] R.L. Rivest, "The RC5 Encryption Algorithm", Dr Dobb's Journal, Vol. 20, No. 1, Jan 1995, pp 146-148.

[10] R.L. Rivest, "The RC5 Encryption Algorithm", K. U. Leuven Workshop on Cryptographic Algorithms, Springer-Verlag, 1995.

[11] Neeraj Khanna et. al, New Symmetric key Cryptographic algorithm using combined bit manipulation and MSA encryption algorithm: NJJSAA symmetric key algorithm, International Conference on Communication Systems and Network Technologies, 2011, pp. 125 - 130.

[12] D. Gollmann and W.G. Chambers,: Clock-controlled shift registers: a review, IEEE Journal on Communications, Vol. 7 , Issue : 4, May 1989, pp. 525 - 533.

[13] P. R. Suri, Sukhvinder Singh Deora: A Cipher based on 3D Array Block Rotation, IJCSNS International Journal of Computer Science and Network Security, Vol.10 No. 2, February 2010, pp. 186-191.

[14] Mohammed M. Alani: Testing Randomness in Ciphertext of Block-Ciphers Using DieHard Tests, IJCSNS International Journal of Computer Science and Network Security, Vol. 10 No. 4, April 2010, pp. 53-57.
[15] S. Dhall, S.K. Pal: Design of a New Block Cipher Based on Conditional Encryption, Seventh International Conference on Information Technology, pages 714, Las Vegas, NV, 12-14 April 2010, pp. 714 - 718.
[16] R.L. Rivest, A. Shamir, and L.M. Adleman, "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems", Communications of the ACM, Vol. 21, No. 2, Feb 1979, pp. 120-126.
[17] A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications, NIST Specifications. SP800-22.

Dr. (Mrs) Pushpa R. Suri is Professor in the Department of Computer Science and Applications at Kurukshetra University, Haryana, India. She has supervised a number of PhD students in the area of Network Security and Cryptograhic techniques. She has also published a number of research papers in National and International Journals and Conference Proceedings. Her areas of interest are Data Structures, Network Security & Cryptography.

Sukhvinder Singh Deora is a Research Scholar at Department of Computer Science & Applications, Kurukshetra University, Kurukshetra, India. He is M.C.A., M.Phil. (Computer Science) and is working as Assistant Professor in N.C. Institute of Computer Sciences, Israna, Panipat, India. To his credit are many prominent papers in the area of data security and issues related to it published in eminent International/National Journals & Conferences. He is a Reviewer to many International Journals of repute and worked as Editor for Indian Society of Information Theory and Applications (ISITA), India. He is having an industry experience of 1.5 years in the areas of Testing, Java, Database design and security issues. His interest areas include Network Security, Theoretical Computer Sciences, Data Structures, S/W Testing and Database Designing. He is also a permanent member of ISITA and many other professional bodies.

IT-infrastructure of university based on cloud computing

Zhyldyz B. Kalpeyeva[1], Akkyz K. Mustafina[2]

[1] Kazakh National Technical University after K.I.Satpayev, Almaty, 050013, Kazakhstan

[2] Department of information technologies, Kazakh National Technical University after K.I.Satpayev, Almaty, 050013, Kazakhstan

Abstract

The article is devoted to the organization of e-education environment of the university based on cloud computing. Cloud services and one of its components - a virtual work place introduce innovations in an education system. The introduction and the use of cloud services at the university will significantly reduce prices of campus machinery maintaining and servicing and will allow to equate in full measure a distance education with a full-time tuition due to a construction of interactive individual paths of training "in any place and at any time", providing high mobility of students.

Keywords: *Cloud Computing, Virtual work place, E-education environment, Virtualization.*

1. Introduction

Over the past few years, the concept of cloud computing and virtualization has become popular in the field of information technologies. Cloud computing has gradually entered in all spheres of society, including an education sphere.

In a modern society training of students, especially in technical educational institutions, is impossible without using the modern Information Technology (IT) of education. But, as a rule, the institutions have a limited budget for updating Computer Park as often as these rapidly changing computing possibilities of modern computers require. It is not a cheap pleasure for educational institutions to contain specialized computer laboratories for various educational and scientific issues [1]. The same picture is observed with the software when universities can not master expenses to keep appropriate informational service of students [2].

An application of the cloud computing could be an economic exit for educational institutions in this situation. Using of cloud computing concepts in educational process will allow reducing costs for purchasing licensed software and expensive computer park with bulk of memory and disks, also the programs, which are used during the lessons, and work results can be stored in a cloud [3].

Cloud computing is a relatively new term in the world of IT-industry, but nevertheless, has become very popular within the past few years. According to Google Trends statistics [4] the interest to the term "cloud computing" has arisen since 2007 and steadily rises.

Several well-known providers of Cloud computing announced services for educational institutions: Google Apps, Microsoft Live@edu, etc. These services can replace or complement the university system functions, such as electronic mail, scheduling plan, instant messengers, creation and storage of personal documents, providing general access to them [5]. In spite of growing interest to cloud computing in the academic community the technology of cloud computing requires a multifaceted study of subjects. The purpose of this research work is to identify benefits and weaknesses of this technology application in the educational institutions of Kazakhstan.

2. The advantages of cloud computing for educational institutions

Many educational institutions have seriously started to think about implementing and using of cloud computing services in the electronic learning environment of the university.

In the works [5-10] are described how it is possible to use effectively cloud computing in students teaching and training. We will generalize and consider some important questions that technology of cloud computing can help to solve in educational institution.

Deterioration of the IT equipment: First of all, universities will reduce costs for the purchase and maintenance of computer equipment. As a rule, the computer equipment morally and physically quickly becomes outdated, not fulfilling the requirements shown by the modern software [11]. Cloud computing technology helps to solve the problems of equipment depreciation, using the weak computers as "thin clients", the information processing occurs on the server side.

Reducing the cost of software: An important argument for the educational institutions in favor of using cloud services, such as Google's «Apps for educational institutions," or «Live@edu» is that the using of cloud providers SaaS solutions is cheaper than providing the necessary services itself [5]. As a matter of fact, costs are missing. There is no need in acquisition and servicing of the corporate equipments and the software for granting of the given services.

Possibility of the virtual classes, laboratories organization: Technology of virtualization and cloud computing allows organizing virtual classes and laboratories, creating necessary quantity of virtual machines (VM) by quantity of students. Virtual laboratories allow students to work with the remote version of the software without their preliminary installation on stationary computers.

Outsourcing: Transfer of some functions of IT Department of universities to the cloud provider. As a rule, educational institutions willingly send e-mail of students to outsourcing cloud providers. E-mail is a service, which migration is relatively easy to realize, because the technology is simple, standardized and demands the minimum adjustment under needs of specific educational institution. Such companies as Microsoft and Google suggest to educational institutions to place free of charge email accounts of students and employees on the servers. Moreover, they can offer mail accounts with the domain name coinciding with a domain name, used by the educational institution. That is rather attractive to an educational institution.

Data backup: As a rule, a sufficient number of methodical materials is collected at university, for example video lectures, multimedia training courses, audio, video files, various student's projects that requires enough disk space. Creation of the backups of teaching materials and transferring them to the cloud solves problems with disk files and also problems of preservations and data recovery in cases of natural disasters [10] and physical malfunctions of campus servers.

Load balancing of servers: Many cloud providers such as Amazon allow to rent computing capacities as needed. This service can be useful for educational institutions to balance the load on their own server park. There are peak loads of servers in learning process, especially during the session and grading students.

Transition to a format «learning anywhere and anytime»: Cloud computing allows students to have access to your personal working environment in regime 24x7x365 irrespective of a territorial arrangement from any accessible devices (PC, laptop, PDA, etc.) if they have access to the Internet.

Important advantage of cloud computing application should be noted - the *decrease of costs on staff.* The number of institution IT staff decreases due to the transfer of services into the cloud. There is no more necessity to improve the knowledge of specialists in the narrow specialized software and advanced training of IT staff.

3. Practice of cloud computing using in KazNTU

This section describes the experience of implementation and using some cloud services in Kazakh national technical university after K.I.Satpayev (KazNTU).

Kazakh national technical university after K.I.Satpayev is the leading technical university and leader of engineering education in Kazakhstan. The university leads own history since 1934. Today KazNTU – one of leading universities of the country, among technical universities occupies 1 place in a rating of Kazakhstan universities. Being a leading technical university of the country KazNTU actively used information technologies in educational process and has a large computer park. At the university operates a large number of basic and applied applications and services the users of which are students, employees of research subdivisions, the administrative and managerial personnel. University dynamically develops and, therefore, the needs in the IT sector constantly grow. Among the important tasks for the university - the consolidation of IT infrastructure, increasing of its reliability and efficiency of resource utilization, decreases in expenses for service of IT-actives. Quickly growing park of computers and a condition in which they were exploited forced to think about reducing of expenses for managing IT, improving of the work reliability of used systems and reducing of reaction time in the case of problems. The cloud computing technologies and virtualization servers allowed achieving it.

Transition to innovative technologies of cloud computing will allow to university to reduce costs to purchase of the licensed software and expensive computer park with bulk of memory and disks, as also the programs which are used during the lessons and results of work can be stored in a cloud. Also carrying over of educational services to "cloud" will allow moving to a format «learning anywhere and anytime».

Like many educational institutions for KazNTU the first step in using of cloud computing was the transfer of students' e-mail support for outsourcing. E-mail - a basic, well standardized service, which can be easily supported from the outside and certainly is not a key for an educational institution. For the organization of corporate

mail KazNTU used "cloud solutions" from Microsoft, which gives free services of corporate e-mail for educational institutions. This e-mail answers for all technical requirements and gives ability to storing email on Microsoft's remote servers in "cloud", instead of on a local computers. It gives possibility to receive mail on mobile devices, possibility of operative connection to system from any place and at any time.
KazNTU is working on creation a private cloud of educational resources. To date, the University has implemented a number of electronic educational services through the portal of the University: e-library, e-learning courses, e-journals.
Each participant of educational process has virtual "private office" and anywhere and anytime has admission to the materials and the data if there is access to the Internet. All data are stored on the party of the server and necessarily results stand out to users.

4. A virtual workplace

As the following stage of introduction of cloud decision paradigm in KazNTU is planned realization of such directions of cloud computing as, an infrastructure as service (Infrastructure as a Service, IaaS), a workplace as service (Workplace as a Service, WaaS). Works on creation of user's virtual workplace (VWP) are conducted, access to which will be carried out both from university, and from an environment, from any device (the home computer, the laptop, netbook, the smart phone, etc.).
The idea of a virtual workplace creation is development of a convenient workplace for each student, containing those services and the software which are necessary for performance educational tasks, and also providing disk space for storage of documents created by student during training.
However, it was necessary to reconsider the existing server equipment of university for implementation of the virtual workplace based on cloud computing technology. KazNTU has begun reorganization of the IT Infrastructure by implementing server platforms virtualization. The computing infrastructure is constructed on the basis of blade-servers IBM BladeCenter HS22V, systems of storage IBM DS3512 and EMC CLARiiON CX300. The network of data storage systems (SAN - Storage Area Network) is constructed on the basis of Fiber Channel switchboards with throughput to 8 Gbit/sec.
The following issue for university was to choose a virtualization platform, comparing possibilities of various virtualization platforms from such developers, as Microsoft, VMware, Citrix, Red Hat. We have stopped the choice on product VMWare vSphere 5.
VMware vSphere is one of virtualization platforms for creating of the cloud infrastructures, representing a complex of services and applications. It provides stable work of applications and possibility to react to various events quickly [12]. A number of services of an infrastructure are responsible for high level of applications and data availability, as well as high level of safety. vSphere helps to accelerate carrying over existing data centers to a computing cloud, allowing subsequently to increase capacities without damage to high school activity, and provides connection to compatible public clouds.
The virtualization technology allows to create virtual servers with various software and to transfer them on a uniform host-server. Thus there are some virtual servers, workstations, IT devices, etc. on a host-server. Virtual servers can take places function separately and independently from each other some, can be moved in a few minutes from one hardware to another, unite in virtual networks or function as uniform knot of processing of the information. This solution has allowed to raise reliability and autonomy of IT Systems, to provide a continuity of business processes of university, to lower power consumption, to raise efficiency of use of computing resources. Thanks to virtualization technology introduction in KazNTU about 25 % of a server infrastructure is already liberated, on 40 % of power consumption is lowered, floor spaces are released from superfluous hardware maintenance, expenses on acquisition of the hardware for new services are cut down. Besides, server virtualization provides access to virtual computers by means of the web interface. Hence, students can be connected to them at home and work in addition. Possibility of remote access creates conditions for distance learning, interest to which increases every year.

5. Conclusions

In the given article authors have tried to describe aspects of application of technology of cloud computing which can be rather attractive and economically beneficial for educational institutions. And also article purpose was to describe the practical experience of using cloud computing for creation of the electronic educational environment of high school. We wanted to demonstrate our own experience in implementing a number of services based on cloud computing.

References

[1] Antipov, O.E., Belov, M.A., Cheremisina, E.N. (2012). A Role of a Virtual Computer Laboratory based on

Technologies of Cloud Calculations in Up-To-Date Computer Education. Distant and virtual education, 1(1), 50-64.
[2] Seydametova, Z.S., Seitvalieva S.N. (2011.). Cloud services in education. Information technologies in education, 9, 105-111.
[3] Mustafina, A.K., Uskenbaeva, R.K., Kalpeyeva, Zh.B.(2011). Cloud Computing and electronic education resources. Bulletin of Kazakh national technical university after Satpaev, 2(84), 3-6.
[4] Google Trends. Retrieved from www.google.com/trends.
[5] Sclater, N. (2010). eLearning in the Cloud. International Journal of Virtual and Personal Learning Environments, 1(1), 10-19. doi: 10.4018/jvple.2010091702.
[6] Sultan, N. (2010). Cloud computing for education: A new dawn? International Journal of Information Management, 30, 109–116.
[7] Singh, G., Singh, H., Sodhi, N.K.(2012). Cloud computing-future solution for educational systems. International Journal of Enterprise Computing and Business Systems, 2 (1). Retrieved from www.ijecbs.com.
[8] Pocatilu, P., Alecu,A., Vetrici, M. (2010). Measuring the Efficiency of Cloud Computing for E-learning Systems. WSEAS Transactions on Computers, 1(9), 42-51.
[9] Erkoç, M.F., Kert, S.B. (2011, June). Cloud Computing For Distributed University Campus: A Prototype Suggestion. Proceedings of the International Conference on Future of Education, Firenze, Turkey.
[10] Khmelevsky, Y., Voytenko, V. (2010). Cloud Computing Infrastructure Prototype for University Education and Research. Proceedings of the 15th Western Canadian Conference on Computing Education. Kelowna, Canada.
[11] Baskakov, A.B., Kuzmichev, V.S., Simanovsky, E.A. (2011). Appliance of virtualization technology and Cloud Computing in Samara state aerospace university. Proceedings of the XVI All-Russian scientific methodical conference “Telematika’2011”, Moscow, Russian Federation.
[12] Junaid Qayyum, Faheem Khan, Muhammad LaL. Setting up of an Open Source based Private Cloud. IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 3, September 2011, 474-479.

Zhyldyz B. Kalpeyeva PhD student, Kazakh National Technical University after K.I. Satpayev. She has published more than 10 scientific works.

Akkyz K. Mustafina PhD, Director of information technologies Department of Kazakh National Technical University after K.I. Satpayev. She has published more than 40 scientific works.

An Effective Genetic Algorithm for Job Shop Scheduling with Fuzzy Degree of Satisfaction

Prof. Dr. Akeela M.Al-Atroshi [1] Dr. Sama T. Azez [2] Baydaa S.Bhnam [3]

[1]**Department of Computer Science, University of Dohuk, Dohuk, +964, Iraq**
[2]**Department of Software Engineering, Mosul University, College of Computer sciences and Mathematics Mosul, Ninawa +964, Iraq**
[3]**Department of Software Engineering, Mosul University, College of Computer sciences and Mathematics Mosul, Ninawa +964, Iraq**

Abstract

The present study suggests a hybrid new fuzzy-genetic algorithm for solving the job shop scheduling problem. Traditional scheduling method does not keep pace with the requirements of the development in the field of manufacturing. Therefore, the current proposed algorithm offers a hybrid intelligent solution between two approaches: genetic algorithm to arrange the jobs randomly, and applied fuzzy logic to build objective function for genetic algorithm. All these are to find optimal degree of satisfaction that achieves optimal chain to schedule a production order using function max(min(satisfaction degree)). The present study includes the modeling of the objective function and adopting a fuzzy logic to solve the issue of scheduling production orders. The matlab is used for Programmable fuzzy Logic, whereas the C++ is used for programming the genetic algorithm with mechanism for linking C++ and matlab. Finally, the algorithm is tested on instances of 10 working procedures (jobs) and 3 machines. The result shows that the hybrid fuzzy-genetic algorithm has been successfully applied to the job shop scheduling problems.

Keywords: Fuzzy Logic, Genetic Algorithm, Job-shop Scheduling, Degree of Satisfaction.

1. Introduction

Scheduling is widely defined as the process of assigning a set of jobs to resources over a period of time . Effective scheduling plays a very important role in today's competitive manufacturing environment. Performance criteria such as machine utilization, manufacturing lead times, inventory costs, meeting due dates, customer satisfaction, and quality of products are all dependent on how efficiently the jobs are scheduled in the system. Hence, it becomes increasingly important to develop effective scheduling approaches that help in achieving the desired objectives[4,7] .

The scheduling and planning a production order have an important role in the manufacturing system. The diversity of products, increased number of orders, the increased number and size of workshops and expansion of factories have made the issue of scheduling production orders more complicated, hence the traditional methods of optimization are unable to solve them [7,10]. Genetic algorithms are stochastic global optimization methods inspired by the biological mechanisms of evolution and heredity, which have been widely used for scheduling problem in recent years [8].

With respect to related studies, one study suggested a genetic algorithm for scheduling a number of jobs on several machines. Each chromosome is divided into two parts: the first is related with machines and the second is for jobs. Fuzzy logic functions are utilized to calculate the degree of completion time and due date [1]. In the same direction [2] proposed two approaches: Jobs Sequencing List Oriented Genetic Algorithm and Operations Machines Coding. The Oriented Genetic Algorithm has been implemented and compared for solving the Job-Shop scheduling problem. Each approach has its own coding, evaluation function, crossovers and mutations applicable in Job-Shop scheduling problem to minimize the workload of the most loaded machine and the total workload of the machines. Jobs Sequencing List Oriented Genetic Algorithm has been found to be the best out of two approaches to minimize the objectives. While [11] improved genetic algorithm to solve the job shop scheduling problems (JSP) through proposing Taguchi-based genetic algorithm. The TBGA combines the powerful global exploration capabilities of conventional genetic algorithm (GA) with the Taguchi method that exploits optimal offspring.

In this research, the job shop problem is considered. The order production contains n jobs J1,..., Jn with given release dates r1,..., rn and due dates d1,...,dn to be scheduled on a set of m machines M1,...,Mm . Each job j=1,…,n consists of a set of operations determined by a process plan that specifies precedence constraints imposed on the operations. Each order contains N products (J1,

J2, ..., Jn) on M of the machines (M1, M2, .. Mm), and there are several probabilities to arrange business Ji and pass it orderly on the machines. Thus, the research domain is very wide and the intervention may be in context of NP-hard problem for the possibility to generate several different sequences of jobs per order.
This research has used genetic algorithm to solve the problem of selecting an optimum arrangement for jobs. But the problem faced by researchers is to build a special objective function for genetic algorithm. On this basis, a fuzzy hybrid degree of satisfaction has been adopted to be used as an objective function of hybrid fuzzy-genetic algorithm which is proposed in this research, taking into consideration that the system handles open number of jobs and machines. The present study aims to: address the problem of compatibility and integration of hybridization for fuzzy logic and genetic algorithm programmatically and philosophically to be as a new and innovative idea to schedule business.
The large-scale projects often face critical circumstances that prevent accomplishment time, so the research depends on max(min (sat. degree)) as an objective function.
The philosophy of the best worst in the design of objective function is adopted for fuzzy-genetic algorithm to achieve the proposed idea and accordingly get optimal scheduling of jobs. The lowest degree of satisfaction in this series is larger than the lowest level of satisfaction of any other chain within the search space .

2. Providing the basic requirements of the proposed algorithm

The proposed system requires providing of several requirements to support the production manager to make decisions about releasing a production order. The requirements are as follows:

1. Preparation of production order file (order.dat):
The structure of order.dat file represents the order details in terms of the product with code, release_date (first r_inf and final r_sub) and the due_date (first d_inf and final d_sub) for each product in the order. The product passes through several machines (operations), so the processing time is required for each product on all the machines in the production line (routing). Even if we assume that the Pi j represents the processing time for the product i on the machine j, where i = 1 ... n and j = 1 ... m and n represents the number of products in the order and m represents the number of machines in the workshop. The three processing times (optimistic p_a most likely p_b and pessimistic p_c) for each m machines determined in order.dat file. So that, the file structure of order.dat to the process of initializing a single product on the two machines is shown in table (1).

Table 1: Structure of order.dat file

Job	*Machine 1*			*Machine ...*			*Machine m*			*Release_Date*		*Due_Date*	
code	pa	pb	pc	pa	pb	pc	Pa	pb	pc	r_inf	r_sub	d_inf	d_sub
a	5	3	2		…		7	4	6	2	4	3	6
…		…			…			…		…		…	

2. Preparations of Proposed Genetic Algorithm Parameters:
Table(2) illustrates the parameters of genetic algorithm, as identified by the researchers.

Table 2: Parameters of proposed GA

Population Size (pop_size)	6
No. Of Generations	20
Chromosome Size	No. of Jobs in the Order = N
Selection Method	Binary Select Method
Crossover Method	Partially Mapped Crossover PMX
Mutation Method	Double simple swapping
Crossover Probability	100%
Mutation Probability	100%

3. Preparations of Fuzzy Logic Parameters:
All Parameters are initialized in C++ language, release_date and due_date are read directly from the file order.dat and all these parameters are transferred to MATLAB. The following equations show how the coefficient of the time accomplishment is calculated [1,9,13] .

$$C_{1,1} = R_1 + P_{1,1} \quad (1)$$

$$C_{i,1} = \max(R_i, C_{i-1,1}) + P_{i,1} \quad ; for \quad i=2,\ldots,n \quad (2)$$

$$C_{1,j} = C_{1,j-1} + P_{1,j} \quad ; for \quad j=2,\ldots,m \quad (3)$$

$$C_{i,j} = \max(C_{i-1,j}, C_{i,j-1}) + P_{i,j} \quad ; for \quad i=2,\ldots,n \; ; for \quad j=2,\ldots,m \quad (4)$$

4. The general outline of the proposed hybrid algorithm:
Genetic algorithm was proposed to schedule production orders (which are programmed using C++) and, choose Fuzzy logic, which is programmed using MATLAB) as an fuzzy objective function to find optimal sequence for an order products. Figure (1) illustrates the general outline of the proposed hybrid algorithm.

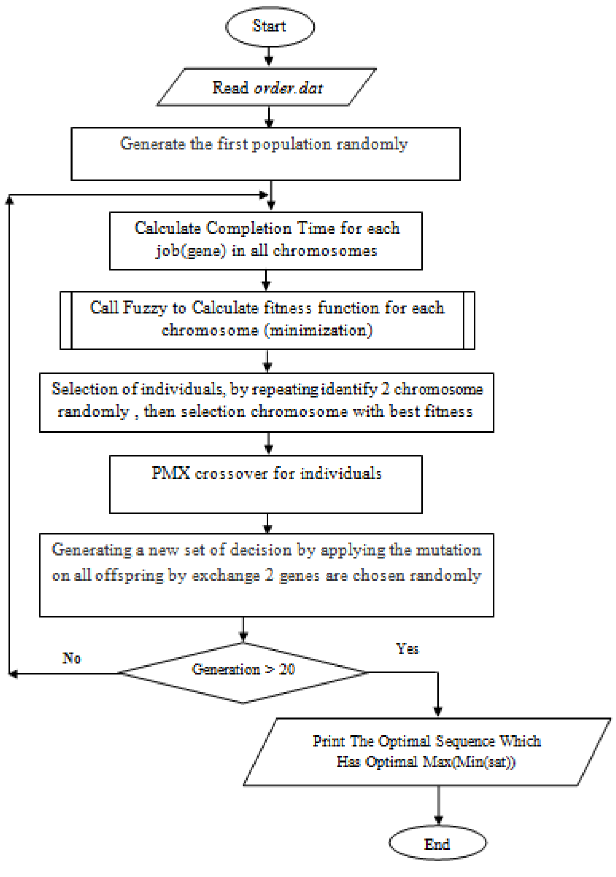


Fig. 1: General flowchart of fuzzy-genetic algorithm

3. Designing a hybrid algorithm and scheduling production orders

Hybrid algorithm was designed using the C++ language , and has been implemented in matlab environment. The implementation strategy is made according to the following sequent steps:

1. Reading the order.dat file: as shown in the table (3), where order of ten jobs is performed on three machines.
No._of_jobs= 10
No._of_machines= 3

Table 3 : Structure and contents of the data file order.dat.

Jobs	Release_date		Process_ time									Due_date	
			M1			M2			M3				
	r-inf	r-sub	Pa	Pc	Pd	Pa	Pc	Pd	Pa	Pc	Pd	d-inf	d-sub
a	2	3	2	3	4	1	2	5	4	5	6	10	15
b	12	16	3	6	7	2	4	5	1	3	4	50	90
c	1	2	5	6	7	3	4	6	7	9	11	11	50
d	6	12	4	6	7	2	5	6	3	6	8	50	80
e	5	9	8	10	11	5	6	7	4	5	8	30	62
f	8	12	9	11	13	7	9	10	2	4	6	41	84
g	4	5	3	6	7	10	12	13	6	7	9	24	78
h	10	13	4	7	9	9	11	13	6	8	9	35	95
i	3	5	8	9	11	10	12	13	5	6	7	18	85
j	3	7	7	9	10	12	13	15	9	11	12	20	70

2. Applying the genetic algorithm: the population size under study is six chromosomes, so the first generation is initialized randomly and ten jobs are ordered randomly. Table (4) shows the first generation as stated in one of implementation cases. Each gene represents the product code in the order file.

Table 4 : The chromosomes of first generation.

1	b	g	d	a	j	f	h	i	c	e
2	j	b	a	i	c	e	h	d	g	f
3	h	e	j	b	d	f	a	i	c	g
4	f	d	a	e	g	b	i	j	c	h
5	a	f	c	j	e	b	g	d	I	h
6	b	a	i	d	e	g	h	j	C	f

3.Callingthe fuzzy algorithm, which is designed to be fuzzy objective function for the proposed genetic algorithm, so, at the same direction, the research used triangular fuzzy release date , due date and comp_time respectively. As shown in figure(2) and (3), fuzzy release date $\widetilde{R}$, denoted by a doublet (rinf , rsup) is described by $\mu_{\tilde{R}}$(rinf)=0 , $\mu_{\tilde{R}}$(rsup)=1. Fuzzy due date $\widetilde{D}$, denoted by a doublet (dinf , dsup) is described by $\mu_{\tilde{D}}$(dinf)=0 , $\mu_{\tilde{D}}$(dsup)=1 [13],and fuzzy comp_time $\widetilde{C}$ denoted by (c1 , c2 , c3) , where $\mu_{\tilde{C}}$(c1)=0,$\mu_{\tilde{C}}$ (c2)=1 and $\mu_{\tilde{C}}$(c3)=0.

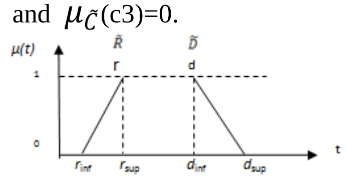


Fig. 2 : Fuzzy release date and fuzzy due date.

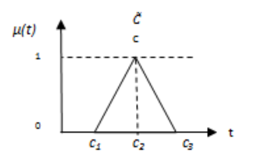


Fig. 3 : Fuzzy completion time.

Three fuzziness criteria are computed and retrieved for each job (gene) in all chromosomes: the objective function represented in a degree of satisfaction SG and the beginning and end time (st , et) respectively is determined as follows:

Repeat for each job (gene) in all chromosomes
if dinf = c2 or dinf = c3 or c2 < dinf < c3
SG=1
St = round (dinf - rsup)/2
et = dinf
else
calculate SG , st and et by :

- Sharing point between $\widetilde{D}$ & $\widetilde{C}$ ($\overline{d\ \ d_{sup}} \perp \overline{c_1\ \ c}$) will determine et on x-axis and SG on y-axis.
- The crossing point between $\widetilde{R}$ and SG ($\overline{r_{inf}\ \ r} \perp SG$) will determine st on x-axis

Endif
End repeat

Table(5) shows the values of fuzziness objective function for the fifth chromosome in the first generation .

Table 5 : Results fuzzy objective function for one chromosome.

jobs	release_date		due_date		Ct			SG	St	et
	r_{inf}	r_{sup}	d_{inf}	d_{sup}	c_1	c_2	c_3			
a	2	3	10	15	9	13	18	0.6667	2.666667	11.666667
f	8	12	41	84	26	36	41	1.00	15.00	41.0
c	1	2	11	50	34	45	52	0.320	1.320	37.52
j	3	7	20	70	50	62	69	0.32258	4.290323	53.870968
e	5	9	30	62	54	67	77	0.1778	5.711111	56.311111
b	12	16	50	90	55	70	81	0.63636	14.545455	64.545456
g	4	5	24	78	64	80	91	0.20	4.200000	67.199997
d	6	11	50	80	67	86	99	0.26531	7.326530	72.040817
i	3	5	18	85	75	96	108	**0.11364**	3.227273	77.38636
h	10	13	35	95	85	109	23	0.11905	10.357142	87.85714

4. Calculating the first part of the objective function min (satisfaction degree),which represents the intermediate results, which are less degree of satisfaction for each chromosome of the current generation. The less degree of satisfaction for each chromosome is stored in crom_mdl_rslt(i), where i= 1,...,6. as shown in table(5). Note that the less degree of satisfaction for the fifth chromosome in the first generation is 0.11364. It is worth mentioning that the SD for second chromosome = 0, so it has less degree of satisfaction among the first generation chromosomes. Therefore, crom_mdl_rslt(2)= 0 and crom_mdl_rslt(5) = 0.11364.
5. Creating the second part of the objective function max(min(satisfaction degree), which represents the final outcome of the optimal degree of satisfaction in the current generation by choosing the chromosome that achieves the highest (best) degree of satisfaction among chromosomes in intermediate results (worst) crom_mdl_rslt. This means selecting the best among the worst max(crom_mdl_rslt) and then storing in the vector gnr_fnl_rslt(i), where i=1,..,20. And, fifth chromosome is the best worst in the first generation and so the gnr_fnl_rslt(1)='afcjebgdih'.
6.Two parents are selected randomly from the current generation, then one of the parents is chosen according to the best degree of satisfaction and this process is repeated until the selection of all parents is done.
7. Crossover process between two consecutive parents to create offspring, on the condition that the gene is not repeated (duplicated) more than once per chromosome by applying the PMX method.
8.Mutation process for all offspring through exchanging two genes chosen randomly.
9.Repeating the previous steps, starting from the third step for the establishment of 20 generations.
10.After the completion of the generating of all 20 populations, finding the best chromosome within vector max(gnr_fnl_rslt(i)), where i= 1,…,20, which is the optimal solution to the scheduling of the order.

4. Implementation of Hybrid Fuzz-genetic algorithm

The hybrid algorithm has been applied to a production order. The details are shown in the table(3). The integrated program for hybrid algorithm (main.cpp) has been implemented and the results as follows:

A: The intermediate results of implementation and the final optimal scheduling result for order.dat are described below:

GENETIC ALG. WITH FUZZY FOR JOB SCHEDULING

No. Of Machines = 3 No. Of jobs = 10

The Minimum Satisfaction Degree Is : 0.098039 for 1st 20 generations

job	Release_Date		Due_Date		Completion_time			Satisfaction_degree	Start	End
a	2	3	10	15	9	13	18	0.666667	2.666667	11.666667
j	3	7	20	70	32	40	44	0.655172	5.620690	37.241379
h	10	13	35	95	38	48	54	0.814286	12.442857	46.142857
c	1	2	11	50	45	57	65	**0.098039**	1.098039	46.176472
b	12	16	50	90	46	60	69	0.814815	15.259259	57.407406
e	5	9	30	62	50	65	77	0.255319	6.021276	53.829788
d	6	11	50	80	53	71	85	0.562500	8.812500	63.125000
i	3	5	18	85	59	78	92	0.302326	3.604651	64.744186
g	4	5	24	78	70	91	104	0.106667	4.106667	72.239998
f	8	12	41	84	73	97	111	0.164179	8.656716	76.940300

(0 Stop / 1 Continue To Generate another 20 Generation). Enter (0 / 1)?: 1

The Minimum Satisfaction Degree Is : 0.113636 for 2nd 20 generations

job	Release_Date		Due_Date		Completion_time			Satisfaction_degree	Start	End
a	2	3	10	15	9	13	18	0.666667	2.666667	11.666667
f	8	12	41	84	26	36	41	**1.000000**	15.000000	41.000000
c	1	2	11	50	34	45	52	0.320000	1.320000	37.520000
j	3	7	20	70	50	62	69	0.322581	4.290323	53.870968
e	5	9	30	62	54	67	77	0.177778	5.711111	56.311111
b	12	16	50	90	55	70	81	0.636364	14.545455	64.545456
g	4	5	24	78	64	80	91	0.200000	4.200000	67.199997
d	6	11	50	80	67	86	99	0.265306	7.326530	72.040817
i	3	5	18	85	75	96	108	**0.113636**	3.227273	**77.386360**
h	10	13	35	95	85	109	123	0.119048	10.357142	87.857140

(0 Stop / 1 Continue To Generate another 20 Generation). Enter (0 / 1)?: 1

The Minimum Satisfaction Degree Is : 0.170455 for 3rd 20 generations

job	Release_Date		Due_Date		Completion_time			Satisfaction_degree	Start	End
a	2	3	10	15	9	13	18	0.666667	2.666667	11.666667
j	3	7	20	70	32	40	44	0.655172	5.620690	37.241379
f	8	12	41	84	34	44	50	0.943396	11.773585	43.433964
c	1	2	11	50	41	53	61	0.176471	1.176471	43.117645
b	12	16	50	90	42	56	65	0.888889	15.555555	54.444443
d	6	11	50	80	45	62	73	0.744681	9.723404	57.659573
e	5	9	30	62	49	67	81	0.260000	6.040000	53.680000
g	4	5	24	78	61	80	91	0.232877	4.232877	65.424660
i	3	5	18	85	70	91	102	**0.170455**	3.340909	73.579544
h	10	13	35	95	80	104	117	0.178571	10.535714	84.285713

(0 Stop / 1 Continue To Generate another 20 Generation). Enter (0 / 1)?: 1

The Minimum Satisfaction Degree Is : 0.203125 for 4th 20 generations

job	Release_Date		Due_Date		Completion_time			Satisfaction_degree	Start	End
a	2	3	10	15	9	13	18	0.666667	2.666667	11.666667
j	3	7	20	70	32	40	44	0.655172	5.620690	37.241379
c	1	2	11	50	39	49	55	0.224490	1.224490	41.244900
h	10	13	35	95	45	57	64	0.694444	12.083333	53.333332
e	5	9	30	62	49	62	72	0.288889	6.155556	52.755554
d	6	11	50	80	52	68	80	0.608696	9.043478	61.739132
g	4	5	24	78	58	75	89	0.281690	4.281690	62.788731
i	3	5	18	85	67	85	97	0.211765	3.423529	70.811768
f	8	12	41	84	71	92	106	**0.203125**	8.812500	75.265625
b	12	16	50	90	72	95	110	0.285714	13.142858	78.571426

(0 Stop / 1 Continue To Generate another 20 Generation). Enter (0 / 1)?: 1

The Minimum Satisfaction Degree Is : **0.265060** for 5th 20 generations **is optimal SD (final result)**

job	Release_Date	Due_Date	Completion_time	Satisfaction_degree	Start	End
a	2 3	10 15	9 13 18	0.666667	2.666667	11.666667
g	4 5	24 78	23 31 36	0.887097	4.887097	30.096775
c	1 2	11 50	30 40 47	0.408163	1.408163	34.081635
j	3 7	20 70	41 52 60	0.475410	4.901639	46.229507
e	5 9	30 62	45 57 68	0.386364	6.545455	49.636364
i	3 5	18 85	52 65 75	0.412500	3.825000	57.362499
d	6 11	50 80	55 71 83	0.543478	8.717391	63.695652
f	8 12	41 84	58 77 90	0.419355	9.677420	65.967743
b	12 16	50 90	59 80 94	0.508197	14.032787	69.672134
h	10 13	35 95	73 96 111	**0.265060**	10.795180	79.096382

(0 Stop / 1 Continue To Generate another 20 Generation). Enter (0 / 1)?: 0

B. Figure (4) and (5) illustrate the Gantt charts for the worst (1st 20 generations) and optimal solution (5th 20 generation) respectively.

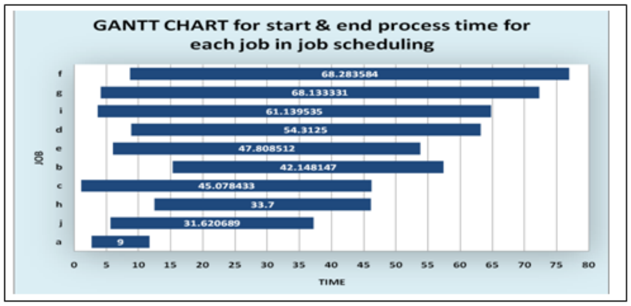


Fig. 4 : Gantt chart for the worst solution of scheduling order.

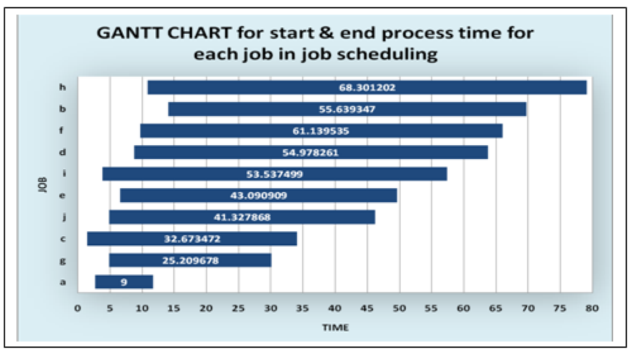


Fig. 5 : Gantt chart for the optimal solution of scheduling order.

C. All the results of implementation begin at the same time start(1'st job)= 2.666667 to all schedules, but the end time of the last job differs from one schedule to another Max(end) as shown in the table(6), which summarizes the result of implementation. Results in the schedule(2) show the highest degree of satisfaction(SD = 1), but the end time of last job in that scheduling is 87.857140 and this situation can be applied to scheduling 3 also, it can be said that it cannot rely Max (SD)as an objective function but its max(end) < max(end) for schedule 2. This undoubtedly supports the hypothesis of research based on a proposal objective function Max (Min (SD)), which is achieved by scheduling5. It is clear that the end time of the last jobs scheduling (time of the completion of the project) is78.571426 and thus has less time to complete the project (for more than 8 hours) compared to scheduling2. Consequently, these results give more flexibility to the technical manager of the project for making a wise decision supported by intelligent techniques on the selection of optimal and best scheduling for the company to complete the project (i.e. the choice between scheduling the first in the case of adopting min(max(end)) and recent in the case of adopting Max (Min (SD)).

Table 6 : A summary of the results of implementation

Running Step	Job Sequence	Min(S.D)	Max(S.D)	Max(end)
1	a j h c b e d i g f	0.098039	0.814815	76.940300
2	a f c j e b g d i h	0.113636	*1.00*	87.857140
3	a j f c b d e g i h	0.170455	*0.943396*	84.285713
4	a j c h e d g i f b	0.203125	0.694444	78.571426
5	**a g c j e i d f b h**	**0.265060**	**0.887097**	**79.096382**

D. The application results have confirmed that the time rate required to complete a single generation is three hundredth of a second (Elapsed time is 0.031 seconds). In other words, the time required to complete the 20 generations and show the best solution is 0.62 or 62 hundredth of a second.

E. The application results have also indicated that the best implementation schedule of the order is obtained through the first ten generations usually. Table(7) shows sequences as various order schedules include their own degree of satisfaction with those obtained during the application. Figure(6) shows optimal solutions chart for twenty generations.

Table 7 : Optimal scheduling for ten generations.

Generation	Optimal Sequence	Min. Sat. Degree
1	a j c h e d g i f b	0.203125
2	a c j e g b i d f h	0.195122
3	a j h c b e d i g f	0.098039
4	a j f c b d e g i h	0.170455
5	g a e i c j f b d h	0.00
6	a c d g e j b i f h	0.216867
7	**a g c j e i d f b h**	**0.265060**
8	a f c j e b g d i h	0.113636
9	a j c b e g d i f h	0.214286
10	a c g j f e i b d h	0.240964

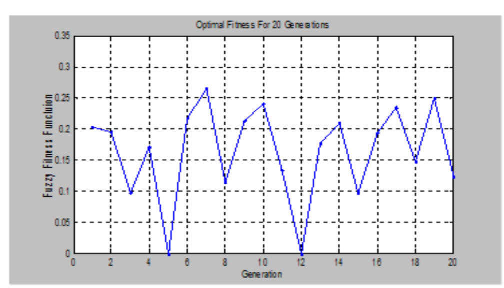


Fig. 6 : optimal solutions chart for twenty generation.

The criteria of fuzziness objective function for optimal solution is (agcjeidfbh) as illustrated in Figure(7).

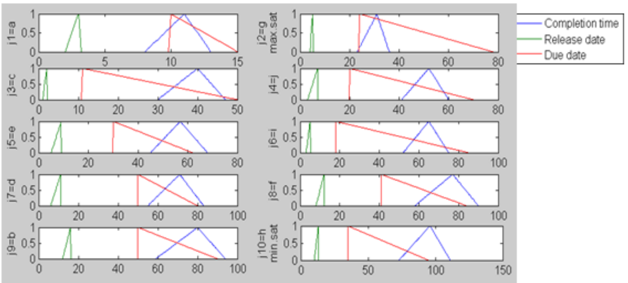


Fig. 7 : The results of the fuzziness objective function for optimal scheduling.

V. Conclusions and Future Works

The designed software has been implemented on an order that contains ten jobs passed on three machines in the job shop to be processed and delivered to the customer. The results were as follows:

1.The proposed system is characterized by specifications of two algorithms: (1)genetic characterized by robustness, reliability and generality [5] as well as (2) fuzzy logic through its ability to handle with incomplete information and sometimes ambiguous[6].

2.The results of the application on several orders confirmed that the best solution can be obtained in the first ten generations of the twentieth generations planned. In other words, best solution can be obtained in the first half of the execution time.

3.There are several traditional ways to resolve the issue of scheduling, but it turns out that the system designer gave special support to resolve the issue of scheduling in a contemporary, technical and intelligent manner, depending on optimization.

4.Hybrid system can be classified as a decision support system, which supports the managers to make their decisions on production scheduling correctly and based on modern and intelligent methods of hybrid.

5.The philosophy of the best worst adopted in the design of objective function genetic algorithm has made the issue of delivery of the order to the customer with the least possible delay in due date.

In the same direction, the current research suggests the concept of scheduling through:

1. Designing a hybrid system through neural networks with genetic algorithm once and with fuzzy logic once again to solve production scheduling problem.
2. Designing a hybrid fuzziness system to integrate the philosophy of the best worst and the philosophy of the highest degree of satisfaction to achieve compatibility between them, and then compared their results with those of the system designer in this current research.
3. Suggesting an objective function min(max(end)) instead of max(min(SD)), especially in the absence of critical conditions for the project.

References

[1] C. Fayad, And S.Petrovic," A Fuzzy Genetic Algorithm for Real-World Job-Shop Scheduling", in M.Ali, F.Esposito, (Eds.), Innovations in Applied Artificial Intelligence,Lecture Notes in Artificial Intelligence 3533, Springer, 2005, pp. 524-533.

[2] A.Kumar1, and R., Singh ," Performance Of Genetic Algorithms For Solving Flexible Job-Shop Scheduling Problem", International Journal of Information Technology and Knowledge Management, vol. 4, No. 1,2011,pp. 105-108.

[3] D.Lei, "Solving fuzzy job shop scheduling problems using random key genetic algorithm", International Journal of Advanced Manufacturing Technology , vol.49,2010,pp.253–262.

[4] H.Metta, "Adaptive, multi-objective job shop scheduling using genetic algorithms", Master's Theses, University of Kentucky,2008.

[5] H.Nanvala," Use of Genetic algorithm based approaches in scheduling of FMS:A Review", International Journal of Engineering Science and Technology (IJEST), ISSN : 0975-5462,Vol. 3 No. 3,2011,pp. 1936- 1942.

[6] H.Nanvala, "Use of Fuzzy Logic Approaches in Scheduling of FMS: A Review", International Journal on Computer Science and Engineering (IJCSE),ISSN:0975-97,Vol.3,No.4, 2011,pp.1734-1739.

[7] Z.Othman,K.Subari, And N.Morad,"Job Shop Scheduling With Alternative Machines Using Genetic Algorithms",Jurnal Teknologi,41(D) , 2007,pp. 67-78.

[8] M.Omar, A.Baharum, Y.Hasan," A job-shop scheduling problem (jssp) using genetic algorithm (GA)". Proceedings of The 2nd IMT-GT Regional Conference on Mathematics, Statistics and Applications. Vol. IV: Computer Sciences & Applications. 13 – 15 June 2006, Penang, Malaysia. pp: 39 – 45.

[9] S.Petrovic, and C.Fayad, "A Genetic Algorithm for Job shop Scheduling with Load Balancing", in AI 2005, Lecture Notes in Artificial Intelligence 3809 edited by S.Zhang and R.Jarvis, pp.339-348.

[10] K.Raja, R. Saravanan, And V.Selladurai, "Mult-Pbjective Optimization Of Parallel Machine Scheduling (PMS) Using Genetic Algorithm And Fuzzy Logic Approach" IE(I)Journal-PR,vol. 87, 2008,pp. 26-31.

[11] J.T. Tsai& others" An improved genetic algorithm for job-shop scheduling problems using Taguchi-based crossover", International Journal of Advanced Manufacturing Technololgy,2008 .

[12] M.Sakawa, and R.Kubota, ,"Fuzzy programming for multiobjective job shop scheduling with fuzzy processing time and fuzzy duedate through genetic algorithms".European Journal of Operations Research, vol.120,2000, pp. 393-407.

[13] X.Song, and S.Petrovic,"Handling fuzzy constraints in flow shop problem", in B.DeBaets, J.Fodor, D.Radojevic (Eds.), Proceedings of the 9th Meeting of the EURO Working group on Fuzzy Sets–EUROFUSE 2005, Belgrade, Serbia & Montenegro, June 15-18,2005, pp. 235-244.

Hybrid Genetic Algorithm and Local Search for Energy Demand Prediction Model

Wahab Musa[1], Ku Ruhana Ku-Mahamud[2] and Azman Yasin[3]

[1] **Electrical Engineering Department, Universitas Negeri Gorontalo Gorontalo, 96125, Indonesia**

[2] **College of Arts and Sciences, Universiti Utara Malaysia Sintok, Kedah, Malaysia**

[3] **College of Arts and Sciences, Universiti Utara Malaysia Sintok, Kedah, Malaysia**

Abstract

Energy demand pattern have many variables related to uncertainty behavior. These lead to a higher estimation rate of energy demand forecasting. However, two problems need to be overcome. The first problem is the fitness evaluation in energy demand forecasting model in which more than one variable are included, and the second problem is the local optimality that single algorithm fails to solve. The objective of this research is to develop energy demand forecasting model that reflects the characteristics of energy demand. A local search is used to assist the genetic algorithm in overcoming uncertainty in demand and the local optima problem and thus producing a higher estimation rate. To evaluate the performance of energy demand model, the actual demand was compared to estimation results. The findings indicate that the solution obtained using the proposed model was an improvement in quality over that obtained by a single genetic algorithm and can be applied to forecast future energy demand with higher approximation accuracy.

Keywords: *Hybrid genetic algorithm, Energy demand forecasting, Higher approximation accuracy.*

1. Introduction

A genetic algorithm is a famous algorithm that has been used in many fields to solve many problems because of its suitability to nearly any function. It simulates the mechanism and the process of evolution, as unique biological features. An algorithm generated from a genetic algorithm, namely the estimation of distribution algorithm (EDA), becomes a hot topic because it is superior. Estimation of distribution algorithms replace some operations in a genetic algorithm, such as learning and sampling of the best individuals of the population, replacing the crossover and the mutation in each iteration of the algorithm [1].

The applications of a genetic algorithm are useful for estimations such as; (i) parameter estimation, (ii) cost estimation, and (iii) energy demand estimation. Parameter estimation using a genetic algorithm has attracted great attention from many researchers. A method for parameter pattern estimation using an adaptive hybrid genetic algorithm was proposed by [2]. The hybrid algorithm is applied to the third-order induction motor. To prevent early convergence in the genetic algorithm, the mutation probability and crossover are changed according to the fitness values of the population at each generation. Their method is successful in solving the problem of parameter estimation in an induction motor.

The study by [3] for parameter estimation of speed governor used a genetic algorithm for optimizing the measurement of frequency and active power variation during transition operation. A real-coded genetic algorithm is applied in all system generators simultaneously to estimate their parameters. It is fully in line to treat and study the comprehensive behavior of a whole electric power system.

An approach for cost estimation has been proposed by [4]. This approach is applying a genetic algorithm to alleviate the drawback of the previous study approach in terms of low prediction accuracy. A previous study proposed effective methods to optimize the weights of the features to estimate the cost with a current project by referring to data collected from past projects. The results of a study by [4] indicated their methods were more effective for software cost estimation than other methods.

The study by [5] for energy demand estimation used a genetic algorithm for optimizing an objective function of electricity demand. When economic growth increases, more needs are created to accompany the higher standard of living, more energy is needed to satisfy energy consumption. The absence of electricity has a negative effect on economic development. During economic boom, a large number of projects for power resources should be constructed. This places heavy pressure on natural resources, the environment and the economy because it is beyond the allowable extent of the national economy [5].

A vital problem in economic development is a study about how to harmonize the fluctuation relationships between electricity construction and economic national development. It also requires scientific demand forecasts for future projections [6]. Such a relationship is a difficult task and some specialists argue that it requires too many inputs and is circular. Causal factors of energy consumption include gross domestic product, oil prices and population growth rate.

Our study proposed a methodology that used population growth, gross domestic product, import, and export as the input variables to determined electricity energy demand as the output of the proposed hybrid genetic algorithm (HGA) for energy demand model. Thus, in the proposed model, energy demand is the function of population, gross domestic product, import and export. Estimation of the relationship between energy demand and independent variables used a HGA is measured by using historical data over a long-term period.

2. Hybrid Genetic Algorithm

There are various applications of hybrid genetic algorithms in solving optimization problems and NP-hard problems. The application of hybrid genetic algorithms can be categorized into three types of hybridization; application of hybrid genetic algorithms with other methods, application of hybrid genetic algorithms with local search and application of hybridization genetic algorithms with parameter adaptation [7].

Genetic algorithms are efficient heuristics and stochastic global search methods that have the ability to handle complicated problems. Unfortunately, these results can only be achieved at the expense of intensive computational requirement. This ability decreases in searching the point that is close to the optimal solution [8]. It can be increased by using a local search capability, which is good at converging at the local optima from nearby starting points. Hybridization of genetic algorithms with other appropriate local search methods would also increase the performance of genetic algorithms in solving global optimization of continuous multimodal functions [9].

Hybridization genetic algorithms with local search are commonly implemented in solving many complex problems where each new generated offspring follows local optimization procedures to lead the solution towards a local optimum area before continuing to the next generation. The local search performs local exploitation around individuals in the local neighborhood, while genetic algorithms make global explorations in a population.

Hybridization of genetic algorithms with local search has proven to provide significant improvement, which was able to explore ability and enhance exploitation towards feasible and highly accurate solutions in solving combinatorial problems [8]; [9].

Therefore, our study aims to apply a hybridization approach in an energy demand pattern forecasting problem and combines local search with genetic algorithms to guide the search towards a feasible solution that minimizes the run time.

3. Proposed HGA Technique

Figure 1 shows the main phases of the proposed hybrid genetic algorithm and the local search algorithm.

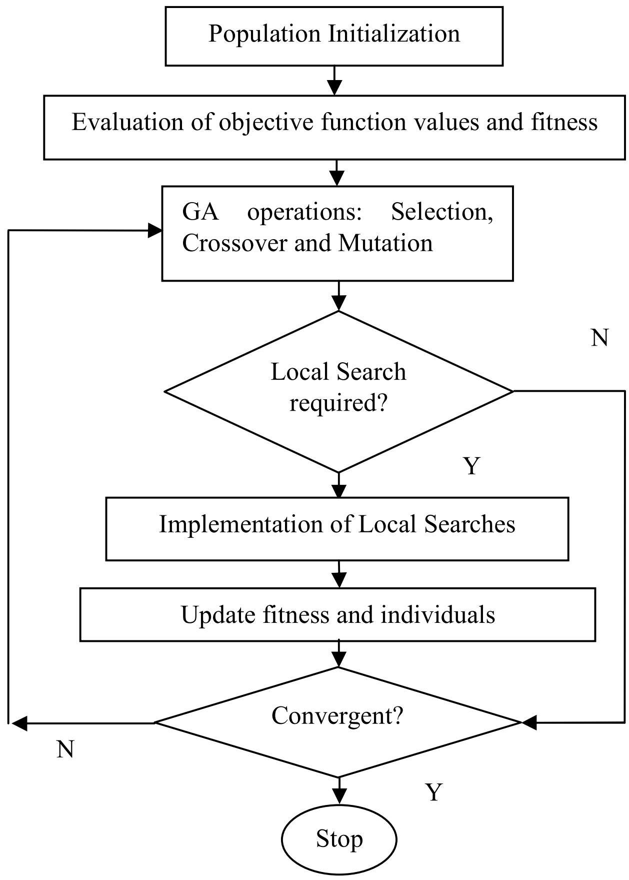


Fig. 1 Proposed HGA Technique

The technical aspect combining a local search algorithm and a genetic algorithm is referred to as the HGA approach. The combining process starts by running GA with small iterations in order to be effective in computational time. During the process, the capability of GA to find a solution quickly in exploring the area of solutions is the main consideration [10].

The process of the proposed HGA starts from the initial population of parameter values. Their objective function values are calculated using an objective function through a genetic algorithm process.

Step – 1. Population Initialization; the population initialization is the process of generating the initial population of *N-pop* chromosomes where *N-pop* is the population size. Heuristic methods and random methods can be used to initiate the population. The commonly used random methods generated $N\text{-}_{Pop}$ x N parameter values between 0 and 1 and then all of these values are scaled to their feasible ranges. However, the heuristic method requires some prior knowledge about the parameter set. One way is take default values as one chromosome and other chromosomes are generated randomly.

Step – 2. Evaluate Objective Function: the model outputs are passed to the objective function to calculate the objective function value associated with the chromosomes, a fitness value is calculated and assigned to each chromosome based on its objective function value.

Step – 3. GA operation: In the execution of a genetic algorithm, one generation is formed by the process of evaluation, selection, crossover and mutation. A new generation of chromosomes is produced after the process is complete. However, there is no guarantee that the best chromosome is carried through to the next generation, the new generation might not be better than the previous one. The genetic algorithm terminates if conditions are satisfied. First, if the objective function value is below the prescribed threshold, the genetic algorithm terminates with an optimal solution. Second, if the maximum number of prescribed generations has been reached, the genetic algorithm terminates without an optimal solution.

Step – 4. Local Search process: The local search process is required based on the objective function values after the genetic algorithm operations. Several options can be taken after the GA operations. If the genetic algorithm process is convergent, then stop: otherwise, return to the GA operations. If a local search is required, carry out the local search. Update fitness and individuals. If the local search process is convergent, then stop: otherwise, continue the GA operations. This process is continuous until convergence is achieved.

4. Performance of Proposed HGAED

Several experiments were done to obtain the appropriate method for solving the energy demand problem. This includes experiments on a single algorithm for energy demand pattern, application of pre-processing and the local search method, and comparison between the proposed HGAED and others method.

Table 1.Comparison of single genetic algorithm, pre-processing and local search

Data	1	2	3	4	5
Average errors (%)					
GA	29.8392	13.369	8.4727	12.5832	8.6511
GA + Prep	25.5609	9.7307	7.2525	9.5239	8.3136
GA+ LS + Prep	6.6571	4.389	3.3004	1.7370	3.464

The experiments in first section are results compared from the single GA with and without pre-processing data, the GA with local search, and the GA with local search and pre-processing data for the energy demand model using the available historical data. The experiments in the second section compared the performance of the proposed HGA model with others models.

This study was to investigate single GA performance by measuring the average error using available data of energy demand, testing the effect of using pre-processing data, testing the effect of using local search when combined with GA and testing the pre-processing data in the HGA. Table 1 and Figure 2 presented the findings from the experiments

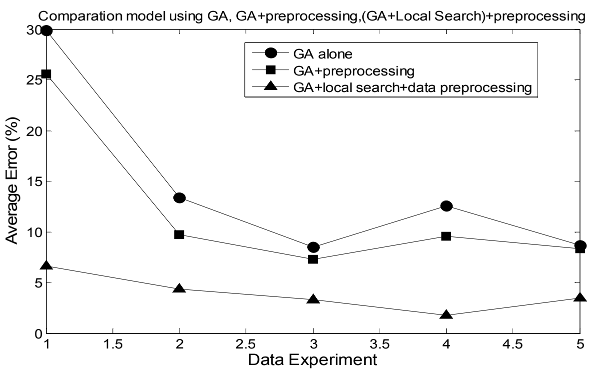


Fig. 2. The average error of single GA and hybrid GA with pre-processing data

Through experimental analysis, results indicated that the hybrid genetic algorithm and local search algorithm using pre-processing data obtained the best solutions. The application of local search combined with genetic algorithms resulted in good solution quality in solving the energy demand pattern problem. Table 2 and figure 3 illustrated the experiments results in the second section. Figure 3 show the performance of the energy demand model, which are measured by the goodness of fit between model outputs and the required target (actual energy demand). It can be seen that the proposed HGAED using the hybrid genetic algorithm and local search approach has better performance.

Table 2. Estimation rate by proposed HGAED and other models

Years	**Actual Demand**	GAED	Linlog	Proposed HGAED
2001	**84.5**	79.521	81.022	75.9079
2002	**87.1**	82.566	76.879	78.7989
2003	**90.4**	88.774	78.544	84.7533
2004	**100.1**	103.651	86.676	92.355
2005	**107.0**	122.129	102.063	105.5662
2006	**112.6**	138.873	116.901	123.0078
2007	**121.2**	155.497	124.667	134.0135
2008	**129.0**	187.313	131.086	137.4572
2009	**136.1**	174.726	111.161	138.4144
Average-Err(%)		**15.293**	**6.426**	**4.389**

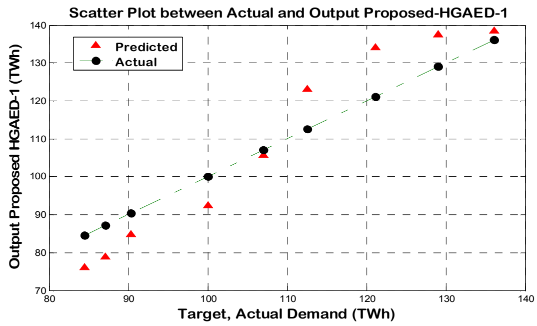


Fig. 3 The goodness of fit to actual demand by proposed HGAED model

5. Conclusions

Based on extensive experiments and obtained results, it appears that the proposed HGAED is more accurate than the conventional genetic algorithm approach. In the proposed HGAED, pre-processing of available data is done before estimation processing, therefore, obtained result proved to have the best accuracy. The performance of the proposed HGAED model was evaluated in terms of error using real energy data with pre-processing. It also was used to predict future demand using a scenario analysis of economic growth.

References

[1] Qiu, Y., Liu, Feng, & Huang, X. "Network Optimization based on Genetic Algorithm and Estimation of Distribution Algorithm". International Conference on Computer Science and Software Engineering, 2008 , Vol. 4, pp. 1058 -1061.
[2] Zhou, X., & Cheng, H. "The induction motor parameter estimation through an adaptive genetic algorithm". 39th International Conference on Universities Power Engineering, 2004, Vol. 1, pp. 494 - 498.
[3] Stefopoulos, G. K., Georgilakis, P. S., Hatziargyriou, N. D., & Sakis Meliopoulos, A. P. "A Genetic Algorithm Solution to the Governor-Turbine Dynamic Model Identification in Multi-Machine Power Systems", 44th IEEE Conference on Decision and Control, 2005, pp. 1288 - 1294.
[4] Li, Y. F., Xie, M., & Goh, T. N. (2007). "A study of genetic algorithm for project selection for analogy based software cost estimation". International Conference on Industrial Engineering and Engineering Management, IEEE, 2007, pp. 1256 -1260.
[5] Ozturk, H. K., & Ceylan, H. "Forecasting total and industrial sector electricity demand based on genetic algorithm approach: Turkey case study". International Journal of Energy Research, No.29, 2005, pp.829-840.
[6]Jian-Chao, H., Zhong-Fu, T., & Xiao-jun, L. "Electricity Consumption and Economic Growth in China: Multivariable Cointegration Analysis and Electricity Demand Forecasting". International Conference on Wireless Communications, Networking and Mobile Computing, 2008. Vol. 8, pp. 1-4.
[7]Asyikin, S.N. "Menu planning model for Malaysian boarding school using self-adaptive hybrid genetic algorithms". PhD Dissertation, College of Arts and Sciences, Universiti Utara Malaysia, Kedah, Malaysia, 2011.
[8] Musa, W., Ku-Mahamud, K.R., & Yasin, A. "Long term energy demand forecasting based on hybrid, optimization: comparative study". International Journal of Soft Computing and Software Engineering, 2012a, Vol. 2, No. 8, pp. 28-36.
[9] Musa, W., Ku-Mahamud, K.R., & Yasin, A. "Hybrid optimization approach to estimate random demand". International Conference on Computer Information Science, 2012b, Vol. 1, pp. 474-479.
[10] Ahmed I, Beg, M.R, Gupta, K.K, & Mansoori, M.I. "A Novel approach of query optimization for genetic population". International Journal of Computer Sciences Issues, Vol. 9, No. 1, 2012, pp. 85-91.

Wahab Musa is a senior lecturer at Electrical Engineering Department, Faculty of Engineering, Universitas Negeri Gorontalo, Indonesia. He received Master of Engineering degree (Electrical Engineering) from ITB Bandung, Indonesia. His research interest includes electricity energy planning, digital signal processing, hybrid genetic algorithm and computational intelligence applied in demand forecasting.

Ku Ruhana Ku-Mahamud is a Professor in the School of Computing at the College of Arts and Sciences, Universiti Utara Malaysia. As an academic, her research interest include computer systems performance modeling, ant colony optimization and computational intelligence.

Azman Yasin is an Associate Professor in the School of Computing at the College of Arts and Sciences, Applied Science Division, Universiti Utara Malaysia, 06010 UUM Sintok, Kedah. Malaysia. His research interest includes software engineering education, information retrieval specifically scheduling and timetabling using artificial intelligence techniques.

Subjectivity Classification using Machine Learning Techniques for Mining Feature-Opinion Pairs from Web Opinion Sources

Ahmad Kamal
Department of Mathematics
Jamia Millia Islamia (A Central University)
New Delhi – 110025, India

Abstract

Due to flourish of the Web 2.0, web opinion sources are rapidly emerging containing precious information useful for both customers and manufactures. Recently, feature based opinion mining techniques are gaining momentum in which customer reviews are processed automatically for mining product features and user opinions expressed over them. However, customer reviews may contain both opinionated and factual sentences. Distillations of factual contents improve mining performance by preventing noisy and irrelevant extraction. In this paper, combination of both supervised machine learning and rule-based approaches are proposed for mining feasible feature-opinion pairs from subjective review sentences. In the first phase of the proposed approach, a supervised machine learning technique is applied for classifying subjective and objective sentences from customer reviews. In the next phase, a rule based method is implemented which applies linguistic and semantic analysis of texts to mine feasible feature-opinion pairs from subjective sentences retained after the first phase. The effectiveness of the proposed methods is established through experimentation over customer reviews on different electronic products.

Keywords: *Subjectivity Classification, Machine Learning, Opinion Mining, Feature Identification.*

1. Introduction

With the exponential growth of World Wide Web and rapid expansion of e-commerce, web opinion sources such as merchant sites, forums, discussion groups and blogs are used as a platform by individual users to share experiences or opinions. Online merchant sites provide space for customers to write feedback about their product and services, as a result number of customer reviews grow rapidly for each product. Such reviews are useful for customers in making purchase decision regarding a product based on the experiences of the existing users, whereas on the other hand, it helps product manufacturers in assessing strength and weaknesses of their products from the perspective of end users. Such information is very useful in developing marketing and product development plans.

Recently, feature based opinion mining technique is gaining momentum in which every granule of customer reviews are processed to identify product features and user opinions expressed over them. However, customer reviews may contain both subjective and objective sentences. Subjective sentences represent user's sentiment, feeling, belief, rants, etc. In contrast, objective contents represent factual information. Consider the following review sentences:

- The battery life of this camera is very *good*.
- Camera is a g*ood* device for capturing photographs.

Both sentences contains opinion bearing word *good*, despite first sentence is subjective and second one is objective in nature. Thus, the target of subjectivity classifications is to restrict unwanted and unnecessary objective texts from further processing. However, classifying a sentence as either subjective or objective is a non-trivial task due to non availability of training dataset. Annotated sets of subjective and objective sentences are difficult to obtain and requires lots of manual processing and thus time consuming [1].

The aim of the current work is to propose methods for identifying subjective sentences from customer reviews for mining product features and user opinions at the intersection of both machine learning and rule-based approaches. In the first phase of the proposed approach, a supervised machine learning technique is applied for subjectivity or objectivity classification for each word of a review sentence, and thereafter the probability of the inscribing sentence to be subjective or objective is calculated using a unigram model. In the next phase, extracted subjective sentences are taken as input by rule-based method which applies linguistic and semantic analysis of texts to identify information components. Initially, an information components are extracted to fill a template $<f, m, o>$, where f represents a product feature, o represents an opinion expressed over f, and m is a modifier used to model the degree of expressiveness of o.

Since, for a product feature, different users may express same or different opinions and a single user with in a review document may express opinions on different features, the basic assumption is that a simple frequency based summarization of the extracted feature-opinion pairs is not suffice to express their reliability. In line with [2], noisy extraction of pairs is handled by calculating *reliability score* for every candidate feature-opinion pair. The value of *reliability score* determines the reliability of an opinion expressed over a product feature.

The remaining paper is structured as follows: section 2 presents a brief review of the existing work in subjectivity classification, product feature and opinion identification. Section 3 presents the architectural and functional details of the proposed system. The experimental setup and results evaluation are presented in section 4. Finally, section 5 concludes the paper.

2. Related Work

The purpose of subjectivity and objectivity classification in opinion mining research is to distinguish between factual and subjective remarks present in customer reviews. Such classification of texts can be performed both at document and sentence levels. The aim of document level subjectivity classification is to identify documents containing subjective texts from large collections for further processing. Due to availability of star rated (1 to 5 stars) customer reviews at merchant sites, divisions among subjective and objective documents are simple. Higher star rated document can be placed in subjective class, whereas lower star rated document can be assigned to objective class [1, 12]. However, a study in [1] revealed that many documents contain combination of both subjective and objective sentences. Subjective texts may also include some factual contents. For example, a movie review usually considered as subjective document (as it reflects sentiments and feelings of its viewers) may contain factual description regarding actors, plot, and list of theaters where the movie is currently playing. On the other hand, objective documents such as newspaper article may enclose subjective texts. Wiebe *et al* in [3] have reported 44% subjective sentences in objective news collection after discarding editorial and review articles. Thus, for better classification performance, sentence level subjective or objective analysis is proposed by various researchers. Many research efforts acknowledge the presence of *adjective* in a sentence as a good indicator for sentence subjectivity. Study in [4] revealed that adjectives are statistically, significantly and positively correlated with subjective sentences in the corpus on the basis of the log-likelihood ratio test. If there exist at least one adjective in the sentence, the probability of a sentence being subjective is 55.8%, despite of more objective than subjective sentences in the corpus. In [5], Hatzivassiloglou & Wiebe study the effects of dynamic adjectives, semantically oriented adjectives, and gradable adjectives on a simple subjectivity classifier and establish that they are strong predictors of subjectivity. Their prediction method for subjectivity states that a sentence is classified as subjective if at least one member of a set of adjective occurs in the sentence otherwise objective. In addition to use *adjective* as a subjectivity indicator in a sentence, study in [6] reported that *noun* can also be used for subjectivity determination. Their experiment using naïve Bayes algorithm achieved a precision of 81% for sentence level subjectivity classification task.

Apart from analyzing subjective or objective sentences in customer reviews, another important task in the field of feature based opinion mining research is to identify product features and user's opinions expressed on them. In [7], Hu and Liu have used an unsupervised method and applied a three-step process for features and opinions extraction. In [8], semi-supervised technique double propagation is proposed to extract features and opinion words using seed opinion lexicon. Further, extracted features and opinions are exploited for identifying new features and opinions. In [9], an unsupervised mutual reinforcement approach is proposed which clusters product features and opinion words simultaneously by fusing both content and link information. Clustering can be very useful for domains where the same feature word is referred by its various synonym. Authors in [10], proposed a supervised approach for movie reviews that apply grammatical rules to identify feature-opinion word pairs. Since, a complete opinion along with its relevant feature is always expressed in one sentence [11], the feature and opinion pair extraction can be performed at sentence level to avoid their false associations.

3. Proposed Subjectivity Classification and Feature-Opinion Pair Mining Method

In this section, architecture and functional details of the proposed subjectivity or objectivity classification and feature-opinion pair mining methods are presented. Fig. 1 presents the complete architecture of the proposed system, which consists of five different functional components – *data crawler, document pre-processor, subjectivity/objectivity analyzer, feature and opinion learner*, and *feasibility analyzer*. Further details about these components are presented in the following subsections.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 1, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

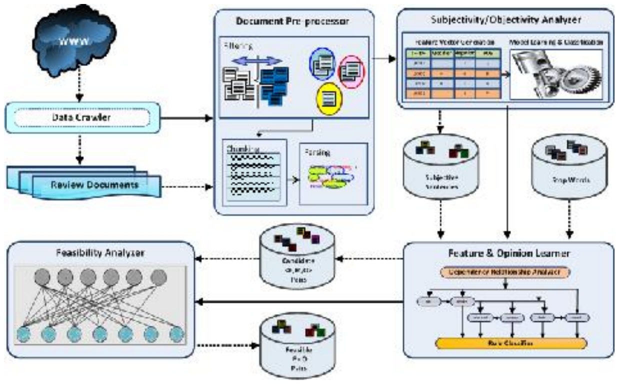


Fig. 1: Architecture of the proposed system

3.1 Data Crawler and Document Pre-processor

For a target review site, the data crawler retrieves review documents and store locally after filtering markup language tags. The filtered review documents are divided into manageable record-size chunks whose boundaries are decided heuristically based on the presence of special characters. For facilitating subjectivity classification and information component extraction, linguistic as well as semantic analysis of text is performed by assigning Parts-of-Speech (POS) tags to every word of a review sentence using POS analyzer. The POS tag reflects the syntactic category of the word and plays vital role in identification of relevant features, opinions and modifiers from review sentences. In this proposed work, POS based filtering mechanism is applied to avoid unwanted texts from further processing.

3.2 Subjectivity/Objectivity Analyzer

Machine learning approaches are likely to provide more accurate classification results, and very useful in learning patterns for identification of subjectivity/objectivity in customer reviews. A supervised machine learning technique is proposed for subjectivity or objectivity classification of each word present in a review sentence, and thereafter the probability of the enclosing sentence to be either subjective or objective is calculated using a unigram model. However, for machine learning application to work effectively, the important task is to engineer set of features and their formulation in a way to produce best classification result. For this purpose, feature vector generator is implemented and attributes such as *term frequency*, *parts of speech*, *opinion indicator seed word*, *position*, *negation*, *and modifier* are used to build a binary classification model for characterization of candidate subjective and objective unigrams from a review sentence. In line with [12], the formulations of different features are presented in the following section.

TF-IDF: It combines the frequency of a unigram in a particular review document with its occurrence in the whole corpus. It is calculated using equation (1), where f is the frequency count of the unigram in the review document, s is the size of the review document in terms of words, N_f is the number of review documents in the corpus containing the unigram, C is the total number of review documents in the corpus.

$$TF-IDF = \frac{f}{s} \times (-\log_2 \frac{N_f}{c}) \qquad (1)$$

Position: It determines the position of the occurrence of candidate unigram in a review sentence. Sometime, position of the unigram plays an important role in deciding sentence subjectivity. The position attribute is set to -1, 0, and 1 in case candidate unigram occurs in the beginning, in-between, and end respectively of the enclosing review sentence.

POS: Part-of-Speech (POS) information is one of the most promising among all features, and used commonly in subjectivity detection. A large number of researches reveal *adjectives* as a good indicator of opinion. Further, nouns (*problem*, *pain*, *issue*, *love*), verbs (*degrade*, *like*) and adjectives (*hard*, *pretty*) can also be used for subjectivity determination of a word. Feature value is set to *A*, *D*, *N*, *V*, and *E* in case candidate unigram *is adjective*, *adverb*, *noun*, *verb*, and *any other* respectively.

Opinion Indicator Seed Word: Opinion indicator seed words are commonly used by reviewers for expressing positive or negative sentiment regarding product features or services and can be used as a good indicator for subjectivity determination. For example, set of positive seed words {*amazing*, *awesome*, *beautiful*, *decent*, *nice*, *excellent*, *good}* and set of negative seed words {*bad*, *bulky*, *expensive*, *faulty*, *horrible*, *poor*, *stupid}*. The feature value is set to 1 in case candidate unigram belongs to positive seed set, and set to 0 if it belongs to negative seed set.

Negation: Presence of negation is also treated as an important clue for subjectivity detection. In case the candidate unigram is a negation word, the feature value is set to 1 otherwise set to 0.

Presence of Modifier: A modifier word usually adverb is used to express the degree of expressiveness of opinion in review sentences. If the candidate unigram is found to be a modifier, then the feature attribute is set to 1 otherwise set to 0.

Class Attribute: This attribute is defined only for the training set of unigrams. If a unigram is subjective, then its value is set to '*S*', otherwise '*O*'.

Table 1 shows a partial list of the candidate subjective and objective unigrams. Table 2 shows an exemplar feature vector to represent candidate subjective/objective unigram of a review sentence.

Table 1: Exemplar subjective and objective unigrams

Subjective unigram	Objective unigram
amazing, beautiful, cheap, decent, effective, fantastic, good, happy, impress, jittery, light, madly, nice, outstanding, perfect, quick, responsive, sharp, terrible, ultimate, wonderful.	access, because, chance, default, entire, few, go, half, inside, job, keep, know, last, matter, new, only, past, quality, read, several, text, use, version, was, young.

Table 2: Feature vectors for subjective/objective unigrams

TF-IDF	Position	POS	Opinion indicator seed word	Negation	Modifier
0.0058	1	*A*	1	0	0
0.0110	0	*D*	0	1	0
0.0232	1	*N*	0	0	0
0.0067	0	*D*	0	0	1
0.0044	1	*E*	0	0	0
0.0412	0	*A*	1	0	0
0.0032	0	*D*	0	1	0
0.0352	-1	*N*	0	0	0
0.0033	0	*D*	0	0	1
0.0062	0	*A*	0	0	0

The proposed method for subjectivity and objectivity determination works in two phases – *model learning* and *classification*. The first phase, also called training phase, uses feature vectors generated from training dataset to learn a classification model, which is later used to identify subjective unigrams in new dataset. The second phase is centered on classification of subjective unigrams from test dataset using the learned model. To determine the subjectivity of a review sentence, it is tokenized into unigrams and the class of each token is determined using the trained model. Finally, the sentence is considered as a subjective sentence if the predicted class for any of token is subjective. For implementation of the classification model, the naïve Bayes algorithm implemented in WEKA [13] is used due to its best performance.

3.3 Feature and Opinion Learner

This module is implemented as a rule-based system, and accepts subjective POS tagged review sentences as input along with dependency relationships information between words. To tackle the peculiarity and complexity of review documents, various rules are defined to access different sentence structures for identification of information components embedded within them. Table 3 represents exemplar review sentences and corresponding dependency relationships information generated by the Stanford parser [14] are shown in table 4.

Table 3: Example review sentences with features, opinions & modifiers

Example Sentence	Feature	Mod-ifier	Opinion
Samsung S5830 has a powerful battery.	Samsung S5830, battery	-	powerful
The picture quality is really nice, amazing and awesome.	picture quality	really	nice, amazing, awesome

Table 4: Example sentences with dependency relationships

Dependency relationships between words
nn(S5830-2, Samsung-1) *nsubj*(has-3, S5830-2) *det*(battery-6, a-4) *amod*(battery-6, powerful-5) *dobj*(has-3, battery-6).
det(quality-3, The-1) *nn*(quality-3, picture-2) *nsubj*(nice-6, quality-3) *aux*(nice-6, is-4) *advmod*(nice-6, really-5) *and*(nice-6, amazing-8) *and*(nice-6, awesome-10).

As observed in [8], existing features can also be used to identify new feature words. In the first sentence mentioned above, the word *S5830* of the product *Samsung S5830* is the nominal subject of the verb *has* and the word *battery* is the direct object of it. Thus, *battery* can be identified as a new feature word. Further, "AMOD" relationship can be used to identify *powerful* as an opinion word. In the second sentence, the bigram *picture quality* is a product feature and can be identified using "NN" tag, whereas the word q*uality* is related to an adjective *nice* with "NSUBJ". Thus, *nice* can be identified as an opinion. Further, multiple opinion words *amazing* and *awesome* present in it can be extracted using one or more occurrence of *and* relationship with the opinion word *nice*. Here, "NN" is a noun compound modifier and "NSUBJ" is a dependency relation used in the Stanford parser. Based on these observations, various rules are designed and reported in [2, 15]. Some sample rules are presented below to highlight the function of the system.

Rule-1: In a dependency relation *R*, If there exist relationships *nn*(w_1 ,w_2) and *nsubj*(w_3 ,w_1) such that POS(w_1)= POS(w_2)= NN*, POS(w_3)=VB* and w_1, w_2 is not a stop-words, Else-If, there exist a relationship *nsubj*(w_3 ,w_4) such that POS(w_3)=VB*, POS(w_4)=NN* and w_4 is not a stop-words, then we search for *dobj*(w_3,w_5) relation. If *dobj* relationship exists such that POS(w_5)=NN* and w_5 is not a stop-words then (w_1, w_2) or w_4 as well as w_5 are assumed as features. Thereafter, the relationship *amod*(w_5 ,w_6) is searched. In case of presence of *amod* relationship, such that POS(w_6)=JJ* and w_6 is not a stop-words, then w_6 is assumed as an opinion.

Rule-2: In a dependency relation *R*, If there exist relationships $nn(w_1, w_2)$ and $nsubj(w_3, w_1)$ such that POS(w_1)= POS(w_2)= NN*, POS(w_3)=JJ* and w_1, w_2 is not a stop-words. Else-If, there exist a relationship $nsubj(w_3, w_4)$ such that POS(w_3)=JJ*, POS(w_4)=NN* and w_3, w_4 is not a stop-words, then either (w_1, w_2) or w_4 is assumed as the feature and w_3 as an opinion respectively. Further, one or more occurrence of $and(w_3, w_k)^+$ is searched where $k \geq 5$. In case of presence of *and* relationship, such that POS(w_k)=JJ* and w_k is not a stop word, then w_k's are identified as opinions.

3.4 Feasibility Analyzer

During the information component extraction phase, various irrelevant nouns, verbs and adjectives are extracted. Sometimes, it is observed that verbs are considered as noun due to parsing error. In line with [2], noisy extractions are handled by calculating reliability score, r_{ij}, for every candidate feature-opinion pair (f_i, o_j), and normalizing this score using *min-max* normalization to scale it in [0, 1] as shown in equation 2, where $HS^n_{(pij)}$ denotes hub score of p_{ij} after n^{th} iteration (after convergence) and *NewMax* and *NewMin* values are set to 1 and 0 respectively. This metric determines the reliability of an opinion expressed over a product feature. Further details about $HS^n_{(pij)}$ can be found in [2].

$$r_{ij} = \frac{HS^n_{(p_{ij})} - \min_{xy}\{HS^n_{(p_{xy})}\}}{\max_{xy}\{HS^n_{(p_{xy})}\} - \min_{xy}\{HS^n_{(p_{xy})}\}} \times (NewMax - NewMin) + NewMin \quad (2)$$

4. Experimental Setup and Evaluation Results

In this section, experimental setup and evaluation results of the proposed system is presented. The data samples used in the experimental work consist of 400 review documents on different electronic product crawled from *www.amazon.com*. The dataset is crawled using *crawler4j API* [1] which is then pre-processed by some filtering to smooth the noise and chunking to decompose the text into individual meaningful chunks or sentences. Using *Stanford Parser API* [2] the text chunks are further broken down to separate the different parts of speech (POS). Standard information retrieval performance measures *precision*, *recall*, and *f-score* are used to evaluate the proposed methods, and defined in equations (3), (4) and (5) respectively.

$$precision = \frac{TP}{TP + FP} \quad (3)$$

$$recall = \frac{TP}{TP + FN} \quad (4)$$

$$f-score = \frac{2 * precision * recall}{precision + recall} \quad (5)$$

4.1 Evaluating Subjectivity/Objectivity Analyzer

A Java based feature vector generators is implemented to generate attributes value for each unigram present in various sentences of the data sample. A total number of 30,000 and 3,800 unigrams are generated from the training and testing datasets, respectively. A binary classification models is made consisting of two classes *subjective* and *objective*. For every unigram generated from a subjective document, the class attribute value is set to *S* otherwise it is set to *O*. From classification results, true positive *TP* (number of correct subjective/objective unigrams the system identifies as correct), false positive *FP* (number of incorrect subjective/objective unigrams the system falsely identifies as correct), and false negatives *FN* (number of correct subjective/objective unigrams the system fails to identify as correct) are obtained. These parameters are used to calculate the value of *precision*, *recall*, and *f-score* using equations (3), (4), and (5) respectively. Further, weighted average *precision*, *recall* and *f-score* values are obtained by considering weight of the two classes used for classification purpose. Weighted average value determines the relative importance of each of the *S* and *O* class on the average result.

4.1.1 Analysis with Feature Attributes

This section is used to discuss the performance of most discriminative feature in the classification task. Table 5 lists the information gain ranking of various features on the basis of WEKA[3] attribute evaluator. POS information is ranked highest i.e. best discriminative features among all for subjectivity/objectivity classification followed by TF-IDF in the experiment. Fig. 2 visualizes the subjective/ objective classification of unigrams on the basis of POS information and TF-IDF using WEKA's visualizer. Majority of the *adjectives* are classified as subjective followed by *adverb*, *noun* and *verb*. Unigrams classified as

[1]http://cod.google.com/p/crawler4j
[2] http://nlp.stanford.edu/software/lex-parser.shtml

[3] http://www.cs.waikato.ac.nz/ml/weka/

subjective are represented by blue colour, and objective unigrams are visible using red colour in fig. 2.

Table 5: Information gain ranking of features

Features	Information Gain
POS	0.10364911
TF-IDF	0.02714459
Negation	0.02082773
Seed	0.00113212
Position	0.00017621
Modifier	0.00000528

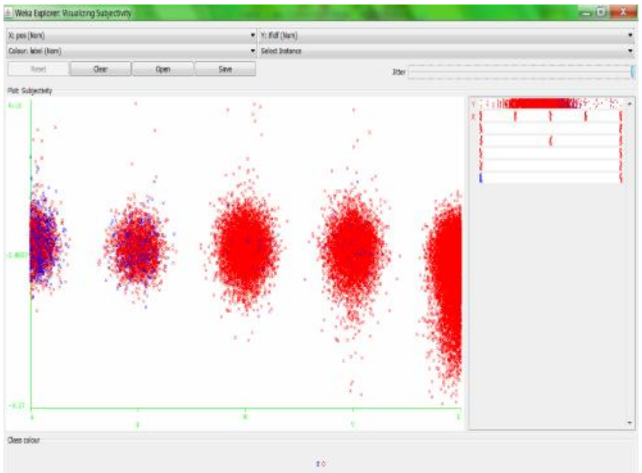


Fig. 2: Visualization of subjective/objective classification of unigrams based on POS information and TF-IDF values

4.1.2 Analysis with Classifiers

Some prominent classifiers are used for experimental purpose best suited for the classification task. Four different classifiers are considered, naive Bayes (a simple probabilistic classifier based on Bayes theorem), J48 (a decision tree based classifier), multilayer perceptron – MLP (a feed forward artificial neural network model with one input layer, one output layer and one or more hidden layers), and Bagging (a bootstrap ensemble method) and 10-fold cross-validation is used for evaluation. For determining real-time applicability of the approach using these classifiers, time consumption by the classifiers is an important concern. Table 6 shows their time consumptions during the experiment. Naive Bayes, being the simplest of all consumes 0.27 seconds, the shortest time duration of all to train the model, whereas MLP takes the longest time of 149.13 seconds. The major demerit of MLP remains in its longer training as well as testing time requirement. Since training needs to be done only once for building the model, longer training time is not a big issue, rather accuracy of the classification task is a major concern.

Table 6: Comparison of time requirements

Classifier	Training Time (in second)	Testing Time (in second)
NB	0.27	0.03
J48	1.12	0.03
MLP	149.13	0.06
Bagging	3.45	0.05

4.1.3 Performance on Training Dataset

As discussed earlier, the training data sample used in the experiment consists of 30,000 unigrams. Fig. 3 shows the summary of correctly and incorrectly instances classified during training by various classifiers used in the experiment.

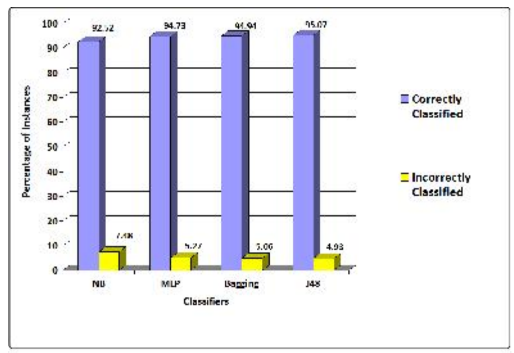


Fig. 3: Classification summary during training

The best classification performance is observed using J48 with correctly classified instances are 95.07%. Naïve Bayes has shown poor performance on training dataset, and percentage of correctly classified instances remains 92.52% only. As discussed earlier, standard information retrieval performance measures are used to evaluate results. In deciding the overall performance of classifiers used in the experiment, *precision*, *recall*, and *f-score* values are obtained for each of the two *subjective* and *objective* classes. As shown in table 7, for subjective class best *precision* (0.722) is obtained using J48, whereas best *recall* (0.668) and *f-score* (0.484) values are emerged from naïve Bayes algorithm. Similarly, for the *objective* class best *precision* (0.981), *recall* (1.0), and *f-score* (0.975) values are retrieved using naïve Bayes, MLP, and J48 respectively.

Table 7: Classifier's performance using IR metrics on training dataset

Classifier	Subjective Class			Objective Class		
	Prec.	Recall	F-score	Prec.	Recall	F-score
NB	0.380	**0.668**	**0.484**	**0.981**	0.939	0.960
J48	**0.722**	0.104	0.182	0.952	0.998	**0.975**
MLP	0.500	0.004	0.009	0.948	**1.000**	0.973
Bagging	0.547	0.229	0.323	0.958	0.989	0.974

Fig. 4 presents ROC curves of all four classifiers, visualizing their comparative accuracy in terms of true positive and false positive rates.

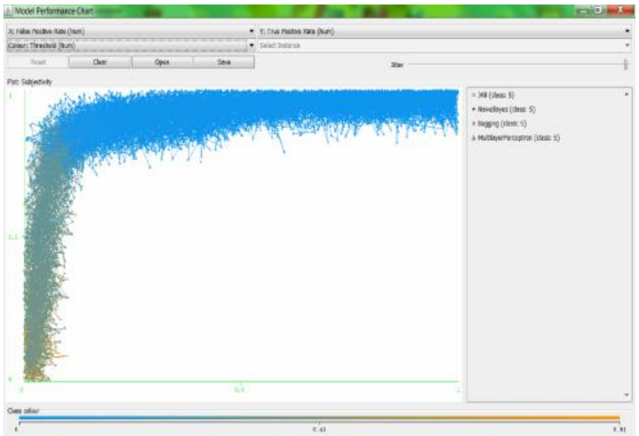


Fig. 4: ROC curves of classifiers for subjectivity/objectivity analysis

4.1.4 Performance on Testing Dataset

Once the training phase of the proposed approach is over, trained model is used to identify subjective or objective unigrams from test dataset. 3,800 instances for testing purpose are framed. Fig. 5 shows the number of instances correctly and incorrectly classified by various classifiers on test dataset. Highest percentage of correctly classified instances i.e. 91.6% is recorded using naïve Bayes followed by J48 with 91.31% accuracy.

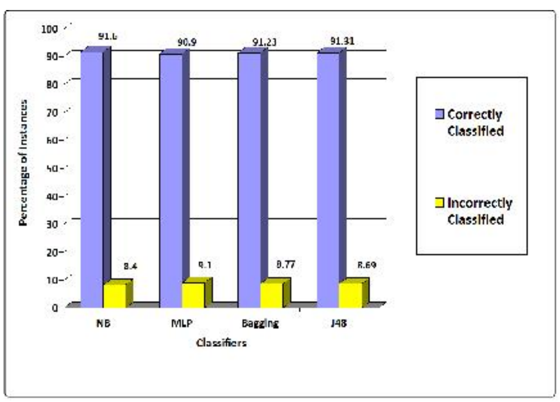


Fig. 5: Classification summary during testing

Similar as training, standard information retrieval performance measures are used to evaluate result on test dataset. For each of the two *subjective* and *objective* classes, *precision*, *recall* and *f-score* values are shown in table 8. For *subjective* class, best *precision* (0.789) is obtained using J48. In symmetry with training results, best *recall* and *f-score* are observed using naïve Bayes with values (0.730) and (0.610) respectively. Thus, for subjective classification, naïve Bayes reflects similar performance with labeled and unlabelled dataset used in the experiment. However, for the *objective* class best *precision* (0.972) and *recall* (0.999) is obtained using naïve Bayes and J48 respectively. However, best *f-score* (0.954) is noted using both J48 and Bagging algorithms.

Table 8: Classifier's performance using IR metrics on testing dataset

Classifier	Subjective Class			Objective Class		
	Prec.	Recall	F-score	Prec.	Recall	F-score
NB	0.523	**0.730**	**0.610**	**0.972**	0.934	0.953
J48	**0.789**	0.044	0.083	0.914	**0.999**	**0.954**
MLP	0.353	0.018	0.034	0.911	0.997	0.952
Bagging	0.554	0.012	0.198	0.919	0.990	**0.954**

In order to obtain information regarding best classifier in hand, weighted average *precision*, *recall*, and *f-score* values are obtained, in which weight (number of instances belonging to a particular class against total number of instances used for the classification purpose) of both *subjective* and *objective* classes are considered. For training dataset, it can be observed from table 9 that the best weighted average *precision* (0.949) and *recall* (0.951) values are obtained using naïve Bayes and J48 respectively. Although, highest weighted average *f-score* (0.939) is achieved using Bagging. However, the performance of naïve Bayes for weighted average *f-score* (0.935) is also comparable due to the next highest average *f-score* value. It is important to note that the better performance of the Bagging method over naïve Bayes is at the cost of the requirement of much more training time.

Table 9: Weighted average values

Classifier	Weighted Average Result (over training dataset)			Weighted Average Result (over testing dataset)		
	Prec.	Recall	F-score	Prec.	Recall	F-score
NB	**0.949**	0.925	0.935	**0.932**	**0.916**	**0.922**
J48	0.940	**0.951**	0.933	0.903	0.913	0.876
MLP	0.924	0.947	0.922	0.861	0.909	0.870
Bagging	0.937	0.949	**0.939**	0.887	0.912	0.886

On testing dataset, best weighted average *precision* (0.932), *recall* (0.916), and *f-score* (0.922) values are maintained by naïve Bayes. Thus, naïve Bayes has emerged as the most suitable classifier in the experiment.

4.2 Evaluating Feature and Opinion Learner

To the best of the knowledge, no benchmark data is available in which features and opinions are marked for electronic products. Therefore, manual evaluation is performed to monitor the overall performance of the proposed system. From the corpus of 400 review documents, a total of 45 documents (Digital Camera: 15, Laptop: 15 and Cell Phone: 15) are randomly selected consisting of 642 sentences for testing purpose. Rule-based method discussed in the section 3.3 is applied to extract feature-opinion pairs. Table 10 presents a partial list of feasible features along with opinions and modifiers.

Table 10: A partial list of extracted features, opinions and modifiers

Product	Feature	Modifier	Opinion
Digital Camera	picture	too, very	glorious, great, excellent, fantastic
	view	really	bad, poor, excellent
	lens	too, quite	good, great, fine
Laptop	sound	pretty, really	great, good, perfect, clear, thin
	weight	extremely	light, noticeable
	price	very, too	higher, great, good, fantastic, reasonable
Cell Phone	player	enough, very	good, nice, great
	screen	pretty, barely, fairly, very	solid, visible, responsive, receptive
	software	rather	easy, slow, flimsy

Initially, the total count obtained for *true positive* (*TP*), *false positive* (*FP*), and *false negative* (*FN*) are 251, 322, and 168 respectively. It has been observed that, direct and strong relationship between words causes extraction of various nouns (or, verbs) and adjectives that are not relevant feature-opinion pairs. As a result, counts for *FP* increase which has an adverse effect on the value of precision. To overcome this problem, a Java based *feasibility analyzer* is implemented to remove noisy feature-opinion pairs. After elimination of noisy pairs, the total count of *FP* reduces to 60. In parallel, manual collection of feature-opinion pairs from test documents are performed. Thereafter, comparing the two sets of pairs *TP*, *FP* and *FN* are calculated. Macro-averaged performance is obtained to present a synthetic measure of performance by simply averaging the result. Table 11 summarizes the performance measure values for the proposed rule-based method in the form of a misclassification matrix. The obtained *recall* (0.599) value is lower than *precision* (0.807), is an indication of system inability to extract certain feature-opinion pairs correctly.

Table 11: Performance evaluation of feature-opinion pairs extraction

Product Category	TP	FP	FN	Precision	Recall	F-Score
Digital Camera	104	23	68	0.818	0.604	0.694
Laptop	70	24	41	0.744	0.630	0.682
Cell Phone	77	13	59	0.855	0.566	0.681
Macro-Average	251	60	168	**0.807**	**0.599**	**0.687**

4.3 Evaluating Feasibility Analyzer

In line with [2], a Java based feasibility analyzer is implemented which compute reliability score for extracted feature-opinion pairs using equation (2). In the beginning of this step, initial score for each feature-opinion pair and review document is set to 1 and the final scores are obtained as soon as convergence of the iterative steps is reached. The convergence is reached when the score computed at two successive iterations for any review document or feature-opinion pair falls below a given threshold i.e. 0.0001. It has been observed that, most of the irrelevant noisy feature-opinion pairs lost their initial hub score, and their final score after convergence reach to a very low value tending towards zero. Table 12 represents a partial list of randomly selected noisy feature-opinion pairs discarded due to very low hub and reliability scores. Table 13 presents hub and reliability scores for some randomly selected feature-opinion pairs from different electronic products. The highest reliability score for pairs *camera-great*, *megapixel-standard*, and *phone-thin* indicates *great*, *standard* and *thin* as the most prominent qualities opined by the reviewers.

Table 12: Noisy feature-opinion pairs with low hub and reliability score values

Feature	Opinion	Initial HS	Final HS (After Convergence)	Reliability Score (r)
screen refreshes	Slow	1.00	0.00	0.00
video recording	bonus	1.00	0.00	0.00
Mcafee	preinstalled	1.00	0.01	0.00
processors	Dual	1.00	0.01	0.00
speedlite	Older	1.00	0.01	0.00

Table 13: Exemplar feature-opinion pairs with hub and reliability scores

Product	Feature	Opinion	Initial HS	Final HS (After Converg-ence)	Reliability Score (r)
Digital Camera	camera	great	1.00	18.11	1.00
	photo	good	1.00	7.76	0.43
	picture	beautiful	1.00	7.16	0.39
	lens	great	1.00	6.30	0.35
	video	good	1.00	5.78	0.32
Laptop	megapixel	standard	1.00	10.59	1.00
	OS	great	1.00	8.48	0.80
	screen	wonderful	1.00	3.43	0.32
	keyboard	great	1.00	3.27	0.31
	price	issue	1.00	2.82	0.27
Cell Phone	phone	thin	1.00	5.70	1.00
	OS	tricky	1.00	2.25	0.39
	screen	large	1.00	1.96	0.34
	camera	good	1.00	1.42	0.25
	keyboard	awesome	1.00	1.07	0.19

In last, table 14 contains top-5 authority scores with their normalized values assigned to review documents of various electronic products.

Table 14: Top-5 authority scored review documents

Product	Authority Name	Initial AS	Final AS (After Convergence)	Normalized AS
Digital Camera	1Canon.txt	1.00	105.13	1.00
	11Kodak.txt	1.00	99.33	0.90
	9Nikon.txt	1.00	96.85	0.86
	21Canon.txt	1.00	95.23	0.83
	13Kodak.txt	1.00	94.93	0.83
Laptop	15Accer.txt	1.00	35.78	1.00
	4Lenovo.txt	1.00	31.60	0.86
	9HpReview.txt	1.00	30.32	0.82
	83Apple.txt	1.00	30.26	0.82
	2Apple.txt	1.00	27.84	0.74
Cell Phone	21LGVuCUC.txt	1.00	27.62	1.00
	7ATTTPhone.txt	1.00	19.85	0.69
	9BlackBerry.txt	1.00	19.63	0.68
	5NokiaNSma.txt	1.00	18.94	0.66
	12LGVuCUC.txt	1.00	18.12	0.62

5. Conclusions

In this paper, the design of a subjectivity/objectivity analysis system is presented based on supervised machine learning approach to identify subjective sentences in review documents. A Java based crawler is implemented to identify opinionated texts and store them locally into record-sized chunks after performing various pre-process steps. Each review sentence is tokenized into unigrams. A set of linguistic and statistical features is identified to represent unigrams as feature vectors and to learn classification models. Various classification models are considered for experimentation to establish the efficacy of the identified features for subjectivity determination. Further, a feature based opinion mining system is presented which implements a rule-based model to identify candidate feature-opinion pairs from subjective review sentences. For every extracted candidate feature-opinion pair feasibility analysis is performed by generating reliability score with respect to the underlying corpus. Standard information retrieval performance measures, including *precision*, *recall*, and *f-score* values are used to measure the accuracy of proposed methods.

References

[1] E. Riloff and J. Wiebe, "Learning Extraction Patterns for Subjective Expressions", Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP-03), 2003, pp. 105-112.

[2] A. Kamal, M. Abulaish and T. Anwar, "Mining Feature-Opinion Pairs and Their Reliability Scores From Web Opinion Sources", Proc. 2nd International Conference on Web Intelligence, Mining and Semantics, ACM, 2012.

[3] J. Wiebe, T. Wilson, and M. Bell, "Identifying Collocations for Recognizing Opinions", Proc. ACL-01 Workshop on Collocation:Computtional Extraction, Analysis, and Exploitation, Toulouse, France, 2001.

[4] R. Bruce and J. Wiebe, "Recognizing Subjectivity:A Case Study of Manual Tagging", Natural Language Engineering, 5(2), 1999, pp. 187-205.

[5] V. Hatzivassiloglou, and JM. Wiebe, "Effects of Adjective Orientation and Gradability on Sentence Subjectivity", Proc. 18th Conference on Computational Linguistics, Association for Computational Linguistics, 2000, pp. 299–305.

[6] E. Riloff, J. Wiebe and T. Wilson, "Learning Subjective Nouns using Extraction Pattern Bootstrapping", Proc. Seventh Conference on Natural Language Learning at HLT-NAACL, Association for Computational Linguistics, 2003, pp. 25–32.

[7] B. Liu and M. Hu, "Mining and Summarizing Customer Reviews", Proc. of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD'04), USA, 2004, pp. 168-177.

[8] G. Qiu, B. Liu, and C. Chen, "Opinion word expansion and target extraction through double propagation", *Association for Computational Linguistics 37*, 2010, pp. 9–27.

[9] Q. Su, X. Xu, H. Guo, Z. Guo, X. Wu, X. Zhang, B. Swen, and Z. Su., "Hidden Sentiment Association in Chinese Web Opinion Mining", Proc. 17th International Conference on World Wide Web, Beijing, China, 2008, pp. 959-968.

[10] L. Zhuang, F. Jing and X.Y. Zhu, "Movie Review Mining and Summarization", Proc. 15th ACM International Conference on Knowledge Management, New York, USA, 2006, pp. 43–50.

[11] B. Li, L. Zhou, S. Feng, and K. Wong, "A Unified Graph Model for Sentence-Based Opinion Retrieval", Proc. 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, Sweden, 2010, pp. 1367–1375.

[12] Bo. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques", Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP), 2002, pp. 79–86.

[13] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, Ian H. Witten, "The WEKA Data Mining Software: An Update" SIGKDD Explorations, Volume 11, Issue 1, 2009, pp. 10-18.

[14] D. M. Marie-Catherine, B. MacCartney, and C. D. Manning, "Generating Typed Dependency Parses from Phrase Structure Parses", Proc. Fifth International Conference on Language Resources and Evaluation, Vol. 6, 2006, pp. 449-454.

[15] A. Kamal and M. Abulaish, "Statistical Features Identification for Sentiment Analysis using Machine Learning Techniques", Proc. International Symposium on Computational and Business Intelligence, IEEE CPS, 2013, pp. 178-181.

Author Ahmad Kamal is teaching Computer Science papers at the Department of Mathematics, Jamia Millia Islamia (A Central University), New Delhi, India. His qualification includes B.Sc. (H) Computer Applications,

Post Graduate Diploma in Bioinformatics (PGDBIN), and Master of Computer Applications (MCA). Currently he is pursuing Ph.D. in Computer Science from the Department of Computer Science, Jamia Millia Islamia, New Delhi. He has more than 6 years of teaching experience at undergraduate and postgraduate levels. His research interests span over the area of *opinion mining*, *natural language processing* and *machine learning*.

Control Theory based Approach for the Improvement of Integrated Business Process Interoperability

Abderrahim Taoudi[1], Bouchaib Bounabat[2] and Badr Elmir[3]

[1] Al-Qualsadi Research & Development Team, ENSIAS, Université Mohammed V Souissi BP 713, Agdal Rabat, Maroc

[2] Al-Qualsadi Research & Development Team, ENSIAS, Université Mohammed V Souissi BP 713, Agdal Rabat, Maroc

[3] Al-Qualsadi Research & Development Team, ENSIAS, Université Mohammed V Souissi BP 713, Agdal Rabat, Maroc

Abstract

Many frameworks are available today to help organizing and performing Enterprise Interoperability projects efficiently. There are also many metrics available to measure the interoperability degree between systems. However, there is a real lack in methodologies to control Enterprise Interoperability improvement projects execution. The aim of this paper is to introduce a new approach to control interoperability improvement projects execution by using control theory, project planning theory and RatIop.

Keywords: *RatIop; Control theory; Enterprise Interoperability; Project planning theory; Interoperability improvement; Automated Business Processes.*

1. Introduction

Interoperability can be defined as the ability for two (or more) systems or components to exchange information and to use the information that has been exchanged [1]. In the current business environment, sharing information and competencies internally, between departments and employees, and externally with partners makes companies much more competitive. A successful implementation of interoperability will help companies to optimize their business processes, reduce their costs, and maximize service quality.

In the Enterprise Interoperability area, many research projects have been launched in the last decades i.e. ATHENA [2], INTEROP [3]. Today, there is a number of frameworks that were developed and validated and are available to use i.e. Chen et al. [4], ATHENA [2], LISI [5], IDEAS [6], EIF [7]. Concerning enterprise Interoperability measurement, many approaches and measures are also available. Ford et al. [8] listed a number of them. There are also other new measures like Chen et al [9] and RatIop [10]. RatIop focuses on measuring the interoperability degree of an automated business process with its environment. It takes into account three main aspects:

- Interoperability maturity level of the environment where the studied process is located.
- Compatibility degree of the external interfaces of the business process with its ecosystem.
- Operational performance of the support systems.

Managing and controlling the execution of interoperability improvement projects raise many challenges. Given the current and targeted interoperability degrees as well as the available resources (i.e. Budget Allocation, Human Resources), the first challenge consists in finding the optimal plan for an efficient management of these projects. The second challenge is the ability to handle unexpected events that can be encountered during project execution, so that the managers can know exactly how many additional resources has to be allocated to correct the deviation from the project optimal plan. The available frameworks and metrics are not currently sufficient to handle the aforementioned challenges.

The aim of this paper is to propose a new approach to Control the execution of interoperability improvement projects. The proposed approach will be based on mature and proven tools: the framework of chen et al [4] (currently under CEN/ISO standardization process) as the interoperability framework, RatIop[10] as the interoperability quantitative metric, Project Planning theory to define the optimal plan and Control theory to control projects execution.

2. Overview of RatIop

RatIop is a new quantitative ratio metric to measure interoperability between automated business processes that was developed in [10]. With this ratio, an organisation can evaluate, at any time and in a quantitative way, the degree of interoperability of its automated business processes. RatIop takes into account three kinds of interoperability measurement as so as:

1. to quantify the first kind of interoperability, Interoperability potentiality, by using the five levels of IMML (Interoperation Maturity Model Level) [10] calculated as bellow:

$$PI = 0.2 * IMML, where\ IMML = 1,2,3,4\ or\ 5 \quad (1)$$

2. to quantify the second kind of interoperability, Interoperability compatibility, by using a modified matrix of Chen et al [10], see Table 1.

Table 1: Interoperability compatibility

	Conceptual		Organizational		Technology	
	syntactic	semantic	Authorities responsabilities	organisation	platform	communication
Business	0/1	0/1	0/1	0/1	0/1	0/1
Process	0/1	0/1	0/1	0/1	0/1	0/1
Service	0/1	0/1	0/1	0/1	0/1	0/1
Data	0/1	0/1	0/1	0/1	0/1	0/1

By noting dc_{ij} the elements of this matrix, this potential is calculated as bellow:

$$DC = 1 - \frac{\sum dc_{ij}}{24}, where\ dc_{ij} = 0\ or\ 1 \quad (2)$$

dc_{ij} is given the value 0 if the criteria in an area marked satisfaction; otherwise if a lot of incompatibilities are met, the value 1 is assigned to dc_{ij}.

3. to quantify the third kind of interoperability, Interoperability performance, by using these three elements:
DS : the overall availability rate of application servers.
QoS : the service quality of different networks used for interacting component communication.
TS : the end users satisfaction level about interoperation.

This potential is:

$$PO = \sqrt[3]{(DS * Qos * TS)} \quad (3)$$

Using these three previous indicators, RatIop is calculated as bellow:

$$RatIop = ((PI + DC + PO))/3 \quad (4)$$

Using this ratio, [11] defines a tool, Interoperability Monitoring Tool (IMT), which has three modules:
Module 1: For assessing interoperability at a specific period.
Module 2: For proposing a scenario to reach a planned degree of interoperability.
Module 3: For giving the prerequisites of going from a maturity level to the next one.

3. Defining the optimal plan of the interoperability improvement projects

Project planning has different meanings in project management. In this paper, Project Planning is the act of building the task by task schedule which we will call the “Project Plan”. The optimal plan is the project plan that minimizes one or more optimization criteria: Cost, Resources and Time. The high level objective of the interoperability improvement projects is to improve interoperability by passing from an initial RatIop R_i, which is the actual state of interoperability, to a targeted RatIop R_t. To define the optimal plan of these projects, we propose to follow these steps:

- Definition of the project objectives
- Definition of the optimal plan using project planning theory.

3.1 Project objectives definition

The high level objective of the interoperability improvement project defined above must be decomposed to clear, concise and measurable objectives which will be used to plan the project properly. To do so, the Periodic Interoperability Monitoring Tool (IMT) [11], can be used to define a clear scenario to reach the desired RatIop Rt. the proposed scenario will define:

- The target Maturity Level.
- The prerequisites to reach this target Maturity Level.
- The incompatibilities to remove.

- The target operational performance ratios: Availability rate of application servers, The QoS of different networks and end users satisfaction level.

3.2 Optimal plan definition

Using the objectives as defined above, there are many planning methods and tools to define the optimal plan taking into account resources, costs and time. The paper [12] lists many deterministic and non deterministic mathematical models used to define optimal plans. Most of these models are already automated. Bellow some examples of these models:

- The standard Project Management model, PMBOK [13].
- Critical Path Method, CPM, and PERT.
- Non-resource-constrained NPV maximization.
- The Resource-Constrained Project Scheduling Problem, RCPSP.
- The Multi-mode Resource-Constrained Project Scheduling Problem, MRCPSP.

The project planning theory will help us define the optimal plan to satisfy the project objectives listed in the section 3.1.
The following table 2 will present a template incorporating the core elements used to define optimal plan:

Table 2: Tasks Description Layout

Task Id	Descri-ption	Precedent tasks	Duration (in weeks)	Resources need	RatIop Elements	RatIop Initial Value	RatIop Target Value

RatIop elements are the elementary components used to calculate RatIop which are: IMML, DS, QoS, TS and the twenty four dc_{ij} (i takes values from 1..4, and j takes values from 1..6).

4. Control of interoperability improvement projects execution

Without careful monitoring and control, many projects fail to achieve the expected results. The aim of this phase is to measure actual execution, compare it with the optimal plan, analyze it and correct the deviations. To achieve this goal, we will use the feedback control theory.

4.1 Feedback control theory

Feedback control theory is widely used in many domains i.e. manufacturing, electronics and physics. It's used also in computer science i.e. apache [14], web servers [15], lotus notes [16], internet [17] and networks [18]. A feedback control system, also known as closed loop control system, is a control mechanism that maintains a desired system output close to a reference using information from measurements of outputs. The feedback control diagram adopted by this paper is illustrated in Figure 1.

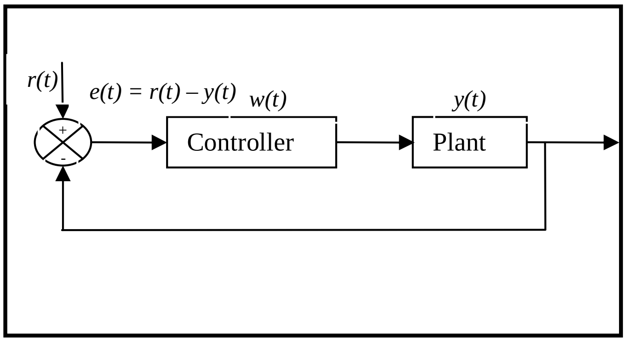


Fig. 1 Feedback control diagram.

The plant is the system to be controlled. In our case, it's the interoperability improvement project. It has a controlled input (denoted by $w(t)$), and a measured output (denoted by $y(t)$). The controller takes as input the control error (denoted by $e(t)$, which is the difference between the observed value and the reference value), and it adjust the input of the plant system to minimize this error. Because of the discrete nature of the system, we will adopt a discrete time approach with uniform interval sizes (Day, Week, two Weeks, or Month).

4.2 RatIop reference Definition

The reference is the RatIop of the system. Its curve will be derived from the optimal plan. We will take into account the finished tasks to calculate the projected RatIop at a time t. The objective of the control system is to minimize the deviations between the desired RatIop based on the optimal plan and the measured RatIop.

4.3 Modeling the plant system

The plant system is the interoperability improvement project. The input of the plant system, at a time t, is the effort consumption at this time to release the project. It can be the resources of the project or budget allocation. The output of the plant system, at a time *t*, is the RatIop at this time. Bellow is the definition of the characteristics of the plant system illustrated in Figure 2:

w(t) = the effort consumption at time t to release the project (resources of the project, budget allocation).

y(t) = measured RatIop of the system at time t

r(t) = the desired RatIop of the system at time t based on the optimal plan.

We will model the plant system as a black-box. We will focus on the behavior of the system not on the internal system construction details which are considered complex. To do so, we will use a statistical approach. The model adopted is the statistical model ARMA. To keep things simple, we will adopt ARMA Model of first order.

$$y(t+1) = a * y(t) + b * w(t) \tag{5}$$

a and *b* are constants which will be estimated statistically. These constants can be estimated by varying inputs (*w(t)*), and calculating the resulting RatIop (*y(t)*). For each value of the effort *w* (resources, budget allocation), an automated project planning software can be used to calculate the optimal plan and derive the values for the RatIop (*y(t)*). Using these experiments, we can estimate the constants *a* and *b* statistically. The use of an ARMA model with greater order will give a more precise approximation of the plant system.

The transfer function of equation (5) is

$$b/(z-a) \tag{6}$$

4.4 Modeling the controller

According to [19], there are four properties of feedback control systems to verify:

- Stability: a system is said to be stable if for any bounded input the output is also bounded.
- Accuracy: a system is accurate if the measured output converges to the reference input.
- Settling time: a system has short settling time if it converges quickly to its steady state value.
- Overshooting: a system that achieves its objectives without overshoot, that is without exceeding an upper limit.

There are three basic controller models:

- Proportional Controller: *w(t) = K*e(t)*
- Integral Controller: *w(t) = w(t-1) + K*e(t)*
- Differential Controller: *w(t) = K*(e(t)-e(t-1))*

The constant *K* is called the gain. To achieve the four properties of our studied feedback control system, the model that we will adopt is the Proportional-Integral model (PI Model):

$$w(t) = w(t-1) + (Kp + Ki) * e(t) - Kp * e(t-1) \tag{7}$$

The transfer function of this PI controller is:

$$Kp + (Ki * z/(z-1)) \tag{8}$$

Thus, we can define the following objectives for our design:

- The system is stable.
- The steady state error is minimized
- The settling time does not exceed a constant value Ks.
- Maximum overshoot does not exceed a constant value Mp.

Using these objectives, [19] discusses in detail the procedure to calculate the appropriate Kp and Ki of the model. With the plant and controller modelled, the control system of interoperability improvement projects is totally defined.

5. Case study

To illustrate the approach, we will use the same e-government example as in [11]. This case consists of an online payment for health care services in a public hospital. It was used in [11] to illustrate RatIop assessment and the usage of the IMT Tool. This system is described in Figure 2.

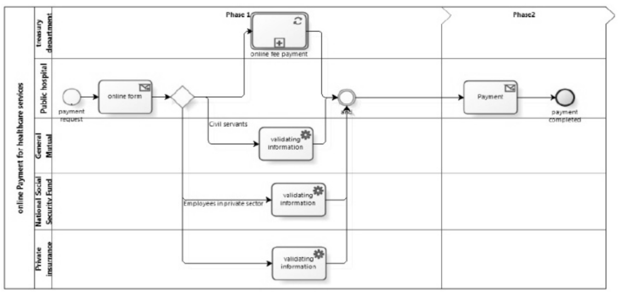


Fig. 2 Online payment business process.

The main objective of this case study is to illustrate the details of steps and calculations used by the approach presented in this paper.

5.1 Initial RatIop assessment

During the implementation phase, three incompatibilities were detected:

- Exchange with mutual servants: Infrastructures are not compatible. It's a Business/Technology platform and communication incompatibility.
- Exchange with National social security fund: Periods for data up-dating not-synchronized. It's a Data/Organizational incompatibility.
- Exchange with private insurance: Process description models can't exchange information. It's a Process/Conceptual syntactic and semantic incompatibility.

The initial interoperability compatibility matrix is listed in Table 3:

Table 3: Interoperability compatibility

	Conceptual		Organizational		Technology	
	syntactic	semantic	Authorities resposabilities	organisation	platform	communication
Business	0	0	0	0	1	1
Process	1	1	0	0	0	0
Service	0	0	0	0	0	0
Data	0	0	0	1	0	0

Using the framework defined in [10] and in section 2 of this paper, the initial interoperability assessment is described in Table 4:

Table 4: Initial RatIop value

Metric	Description	Value
Maturity Level	IMML	0,4
Interoperability compatibility	DC	0,79 (Based on Table 3)
Overall application servers availability	DS	0,9
Network quality of service	QoS	1
End user satisfaction	TS	0,8
RatIop metric	**RatIop**	**0,69**

5.2 Project objectives definition

The targeted RatIop is 0,8. The proposed scenario to reach this targeted RatIop is:

- Remove the tree incompatibilities of the system.
- Improve the Overall application servers' availability to be 1.
- Improve the end user satisfaction level to be 1.

5.3 Optimal Plan Definition

Using these objectives, the project tasks are defined in Table 5. The duration unit is the week:

Table 5: Tasks description

Task Id	Description	Precedent tasks	Duration (in weeks)	Resources need	RatIop Elements	Initial Value	Target Value
Task 1	Removing exchange with mutual servants incompatibilities		3	5	dc_{15}, dc_{16}	0	1
Task 2	Removing exchange with National social security fund incompatibilities		5	6	dc_{44}	0	1
Task 3	Removing exchange with private insurance incompatibilities		5	6	dc_{21}, dc_{22}	0	1
ask4	Improving the Overall application servers' availability		2	2	DS (From 0,9 to 1)	0,9	1
Task 5	Improving the end user satisfaction level		3	5	TS (From 0,8 to 1)	0,8	1

All these tasks are independent. The total resources for the project are 6. The optimal plan is described in figure 3.

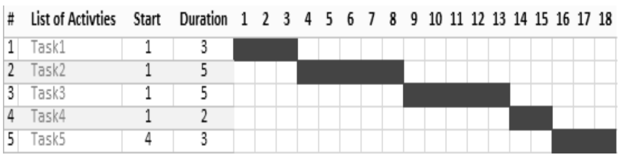


Fig. 3 Optimal plan

5.4 RatIop reference Definition

Using this optimal plan, the RatIop reference is described in figure 4:

Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
RatIop	0,69	0,69	0,69	0,72	0,72	0,72	0,72	0,72	0,74	0,74	0,74	0,74	0,74	0,76	0,76	0,77	0,77	0,77	0,80

Fig. 4 RatIop Reference

5.5 Modeling the plant system

Figure 5 illustrate the evolution or RatIop depending on the resources.

Ressources	Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
6	RatIop*100	69	69	69	72	72	72	72	72	74	74	74	74	74	76	76	77	77	77	80
7	RatIop*100	69	69	71	73	73	73	73	73	76	76	76	76	76	77	77	77	80	80	80
8	RatIop*100	69	69	71	73	73	73	73	73	76	76	76	76	76	77	77	77	80	80	80
9	RatIop*100	69	69	71	73	73	73	73	73	76	76	76	76	76	77	77	77	80	80	80
10	RatIop*100	69	69	71	73	73	76	76	76	76	76	77	77	77	77	77	80	80	80	80
11	RatIop*100	69	69	71	73	73	76	76	76	77	77	77	77	77	80	80	80	80	80	80
12	RatIop*100	69	69	71	76	76	76	76	76	80	80	80	80	80	80	80	80	80	80	80
13	RatIop*100	69	69	71	76	76	76	76	76	80	80	80	80	80	80	80	80	80	80	80
14	RatIop*100	69	69	71	76	76	76	76	76	80	80	80	80	80	80	80	80	80	80	80
15	RatIop*100	69	69	71	76	76	76	76	76	80	80	80	80	80	80	80	80	80	80	80
16	RatIop*100	69	69	71	76	76	76	76	77	80	80	80	80	80	80	80	80	80	80	80

Fig 5 RatIop over time and resources

Using the least square regression method, the plant system parameters estimation is:

$$y(t+1) = y(t) + 0.1 * w(t) \tag{9}$$

5.5 Modeling the controller

The objectives of our design are:

- The system is stable
- The steady state error is minimized
- The settling time *Ks* does not exceed a constant value 20
- Maximum overshoot *Mp* does not exceed a constant value 20%.

Using these objectives, [19] discusses in detail the procedure to calculate the appropriate *Kp* and *Ki* of the model. In our case, the steps followed are:

Step 1: Calculate *r* and θ using the following equations

$$r = e^{\wedge}(-4/Ks) \tag{10}$$

$$\theta = \pi(\log(r)/\log(Mp)) \tag{11}$$

In our case, Ks=20 and Mp=0,2

So: r= 0,819 and θ= 0,39

Step 2: Calculate the desired characteristic polynomial using the following equation:

$$x^2 - 2r\cos(\theta)x + r^2 \tag{12}$$

In our case, the characteristic polynomial is:

$$x^2 - 1{,}51x + 0{,}67 \tag{13}$$

Step 3: Construct and expand the modelled characteristic polynomial
The modelled characteristic polynomial is

$$(K(z) * G(z))/(1 + K(z) * G(z)) \tag{14}$$

Where

$$K(z) = ((Kp + Ki) * z - Kp)/(z-1) \tag{15}$$

K(z) is the transfer function of the PI Controller in equation (8).
G(z) is the transfer function in equation (5)

$$G(z) = b/(z-a) \tag{16}$$

Expending (14) and eliminating all fractions in the denominator will give us the following polynomial:

$$z^2 + [\mathrm{b(Kp + Ki) - 1 - a}]\mathrm{z} + \mathrm{a - bKp} \tag{17}$$

Step 4: Solve Kp and Ki by resolving the equation (12) = (17).
So

$$-2r\cos(\theta) = [\mathrm{b(Kp + Ki) - 1 - a}] \tag{18}$$

And

$$r^2 = \mathrm{a - bKp} \tag{19}$$

In our case: a=1, b=0,1 , r= 0,819 and θ= 0,39
Resolving these two equations will give us:

Kp=3,3 and Ki=1,56

So our controller is modelled as:

$$w(t) = w(t-1) + 4{,}86 * e(t) - 3{,}3 * e(t-1) \tag{20}$$

We can see that the value “4,86” is approximately the mean value of task resources. If the RatIop is less than the reference, the controller will suggest adding this quantity of resources to begin a pending task. This will accelerate the advancement of the project. The proposed approach will be more efficient if these conditions are met:

- Projects are medium to large (more than 50 tasks).
- Choosing the unit of time the largest possible.
- In the plant model, choosing an ARMA model with greater order.

6. Conclusion and Future work

In this paper, we have proposed a complete approach to control the execution of interoperability improvement projects. It’s based on proved mathematical models (feedback control theory and statistics) in addition to the metric RatIop. We have modeled the interoperability improvement project as a black box system without entering deeply into the relationship between input (i.e. work effort) and output (RatIop). In future work, we will try to model the system in more details. We will work also on the applicability of other branches of control theory, like optimal control.

References

[1] IEEE. (1990). IEEE standard computer dictionary: a compilation of IEEE standard computer glossaries.
[2] ATHENA (2003), Advanced Technologies for Interoperability of Heterogeneous Enterprise Networks and their Applications, FP6-2002-IST-1, Integrated Project.
[3] INTEROP (2003), Interoperability Research for Networked Enterprises Applications and Software, network of excellence.
[4] Chen, D., Daclin, N. (2006). Framework for enterprise interoperability. In IFAC TC5.3 workshop EI2N, Bordeaux, France.
[5] C4ISR (1998), Architecture Working Group (AWG), Levels of Information Systems Interoperability (LISI).
[6] IDEAS (2003), IDEAS Project Deliverables (WP1-WP7), Public reports, Retrieved from www.ideas-roadmap.net.
[7] EIF (2004), European Interoperability Framework, White Paper, Brussels, http://www.comptia.org.
[8] Ford, T. C., Colomb, J., Grahamr, S. R., Jacques, D. R. (2007, June). A survey on interoperability measurement. In Proceedings of 12th International Command and Control Research and Technology Symposium, Newport, RI.
[9] Chen, D., Vallespir, B., Daclin, N. (2008). An approach for enterprise interoperability measurement. In Proceeding of MoDISE-EUS. France.
[10] Elmir B., Bounabat, B. (2011). A Novel Approach for Periodic Assessment of Business Process Interoperability. IJCSI International Journal of Computer Science Issues, 8(4), ISSN (Online): 1694-0814, www.IJCSI.org.
[11] Elmir, B., Alrajeh N.A., Bounabat, B. (2011). Interoperability monitoring for e-government service delivery based on enterprise architecture. In International Conference on Information Management and Evaluation (ICIME), Toronto, Canada.
[12] WILLIAMS, T.M. (2003). The contribution of mathematical modelling to the practice of project management. IMA Journal of Management Mathematics 14, 3–30.
[13] Project Management Institute (2000). A Guide to the Project Management Body of Knowledge (PMBOK). Project Management Institute, Upper Darby, PA, US.
[14] Gandhi, N. Tilbury, D. M., Diao, Y., Hellerstein, J., Parekh, S. (2002). Mimo control of an apache web server: Modeling and controller design. In Proceedings of the American Control Conference.
[15] Lu, C., Abdelzaher, T., Stankovic, J., Son, S. (2001). A Feedback Control Approach for Guaranteeing Relative Delays in Web Servers. In IEEE Real-Time Technology and Applications Symposium, Taipei, Taiwan.
[16] Gandhi, N., Parekh, S., Hellerstein, J., Tilbury, D. M. (2001). Feedback Control of a Lotus Notes Server: Modeling and Control Design. In American Control Conference, Arlington, VA, USA.
[17] Mascolo, S. (1999). Classical Control Theory for Congestion Avoidance in High-speed Internet. In Proceedings of the 38th Conference on Decision & Control, Phoenix, Arizona, US.
[18] Chiu, D., Jain, R. (1989). Analysis of the Increase and Decrease Algorithms for Congestion Avoidance in Computer Networks. Computer Networks and ISDN Systems, 17(1).
[19] Hellerstein, J. L., Diao, Y., Parekh, S., Tilbury, D. M. (2004). Feedback Control of Computing Systems. ISBN 0-471-26637-X, John Wiley & Sons.

Service Orchestration Algorithm for Web Services: Evaluation and Analysis

Yousra Chtouki[1, 2], Hamid Harroud[1]

[1]Al Akhawayn University in Ifrane
P.O Box 104, Ave Hassan II
Ifrane 53000, Morocco

Mohammed khalidi Idrissi[2] ,Samir Bennani[2]

[2]Ecole Mohammadia d'Ingénieur
RIME Laboratory; LRIE EMI
Ave Ibn Sina, Rabat, Morocco

Abstract

With the growing need for service oriented architecture and the rise of cloud computing, web services have become the focus of many researchers and software developers. Our area of focus is web service composition. In this paper we are presenting an evaluation and analysis of the proposed orchestration algorithm in [1]. The algorithm is intended to orchestrate interactions between different web services. Generate a workflow of a set of web services in order to generate a composed service. We present an approach that aims to optimize the process of web service composition by assigning the best service and service activities. This involves taking into consideration several parameters such as the execution time, cost of communication between services that may reside in different locations, constraints imposed by the user and the quality of service (eg: response time, cost, reliability, availability, etc..).Our work can be implemented using a set of web services standards therefore it can be integrated in different environments. We are using WSDL for web service description, PBEL for business process description and creation, SOAP, messaging protocol, and WS-Policy to describe and create rules to be applicable on web services orchestration.

KeyWords

Web Services orchestration, web service composition, WSDL, PBEL, SOAP, WS-Policy.

1 Introduction

Web services are described as any application or part of an application that is available over the web. Web services are self-describing and self-contained network available software modules that perform concrete tasks and are deployed easily because they are based on common industry standards and existing technology such as HTML and HTTP[2]. Most web based services are compatible with all popular browsers. Therefore, the issues of software and hardware compatibility is rare if not absent. All of this has pushed software developer and users to opt for web based

services. The challenge is that most existing services are not suited for every user's needs, most of the time users may have to work on multiple services to complete a single task.
In education particularly we have surveyed a group of university professors to find out the web based applications (web services) that they use in daily basis to complete their teaching related tasks. We have found that more that 80% use at least one web service along with the existing learning management system (LMS) We have also found that most of the web services used by educators are general purpose services, most of them are free web based services, easy to use and tend to be very popular services among a large number of different users not necessarily related to education. According to the center for learning & performance technologies [3] the top 10 eLearning tools for 2012 are twitter, YouTube, Google Docs, Google Search, WordPress, Dropbox, Skype, PowerPoint, Facebook, Wikipedia. Surprisingly Moodle which is a learning management system comes as 11th in the list. This further confirms that educators are using services provided by different external web services besides their existing LMS in order to complete daily teaching related tasks. New Web 2.0 technologies and websites such as a blog, wiki or YouTube make new demands on learning, and they provide new supports to learning [4], In a previous publication we have specifically evaluated the impact of YouTube videos on the student's learning [5]. We have found that the use of YouTube videos as a support material to further illustrate concepts and enforce the students learning is very effective. We have also found that student's interest in the subject has increased. From the student's perspective using such tools as YouTube has become necessary because of the change in the nature of the students and the way they learn. They prefer random "on-demand" access to media; expect to be in constant communication with their friends and ease of access in the creation of their own new media [4]. Therefore, in any given educational environment, there is a need to use external web services to accommodate needs not offered by the existing LMS. From this comes the need for web service composition. This is the best way to benefit from different services by different providers without any major changes to existing LMS or any work environment in general.
Since a given web service offers several functionalities, therefore it contains several sub-services. Service composition allows us to use sub-services of different WS and compose them according to the user's requirements: Figures 1a and 1b show the list of web services that may participate in the composition process; we notice that each service is composed of several basic services. Therefore functionalities of existing web services can be used as a basic service. Delicious for instance contains posts and tags as subservices and each one contains other sub-services as well. The possibility to consume part of the web services relieves some of the overhead during the execution time. It also makes the composed service specific to the user requirements.

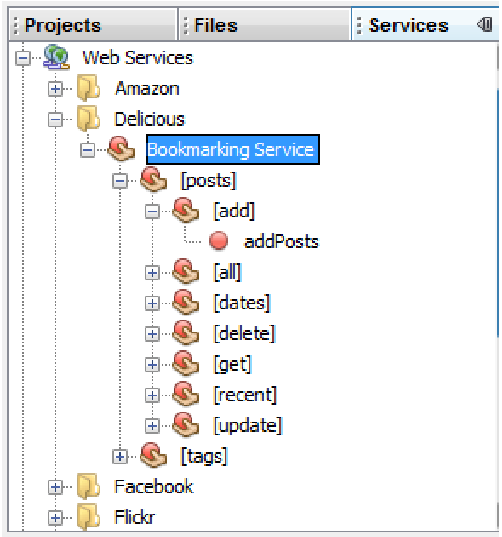


Figure 1a

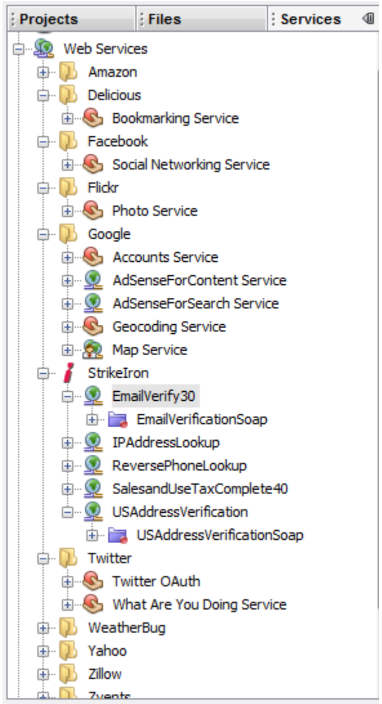


Figure 1b

2 Motivational Example:

We have established that there is an obvious need to integrate multiple web services in order to accomplish daily eLearning tasks. Based on the list of the most popular eLearning tools we opted to pick some of the most popular web services and generate a motivational example based on our teaching practices and needs. YouTube is increasingly being used by educators as a pedagogic resource for everything from newsworthy events from around the world to “slice-of-life” videos used to teach students within an ESL (English as a Second Language) course. From instructional videos to an online space to share student authored content [4]. Thus, the goal is to use YouTube as a source of videos to serve as a learning content and deliver it through the currently used LMS. This requires several steps, first, search for a video by topic then select one, download the selected video this can be done using RealOne downloader plugin. Next, upload the video on the LMS. To complete these steps we need three service providers: YouTube, RealOne, and the LMS. Obviously, each of the participating services has its own rules, restrictions and limitations that have to be taken in consideration during the composition process. This will ensure that the interactions between these services will be smooth and invisible to the end user.

Figure 2 outlines the video request process, in which the user submits the request and the goal is to make use of all three services LMS, YouTube and RealOne. The user submits the request to search for a video and upload it to the LMS, the request is submitted through the orchestrator plugin that can be plugged to any web application. In this case the request is submitted through the LMS. Since the LMS doesn’t have a video search the search for video is sent to YouTube (the video WS). YouTube executes the video search request, returns a list of matches to the user, the user makes a selection. Then the YouTube service checks for the RealOne Downloader plugin, if it's found the download of the selected video is executed and the file path is saved and returned as output of YouTube WS.

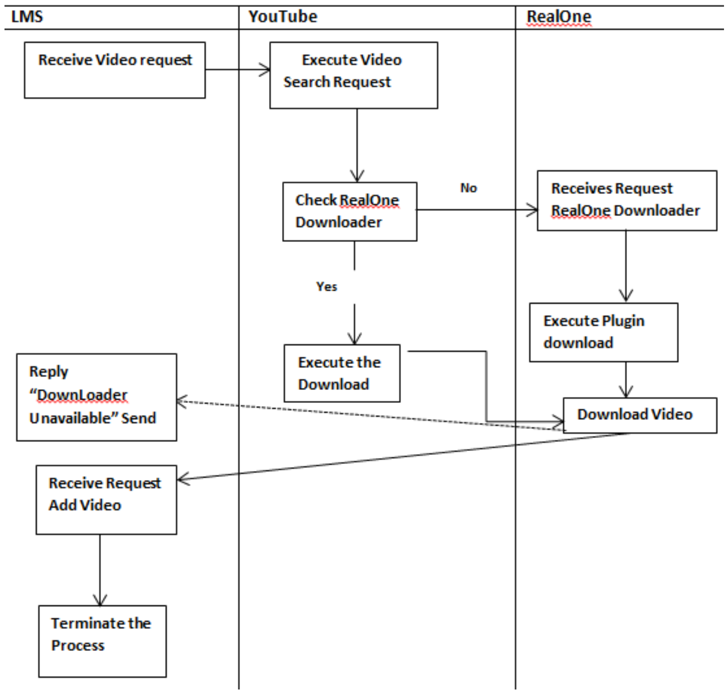


Figure2: Outline of a video request process

3 Service Composition:

Service composition is a very complex process, it includes several levels: The basic services, their description and how they will be integrated and triggered by the system. The orchestration and composition process, which includes finding the right set of services; generate a workflow and determine the interactions between the participating services; finally, the composed service and how it will be presented, stored and reused for future compositions.

The goal of our work is to find an orchestration algorithm that will provide optimum results in terms of managing interactions between the different participating services. Next, ensure that the rules specific to each service are applied whenever that service is involved in a composition process. Then, enforce rules specific to the interactions between the participating services. Since web based applications can be developed using different languages and have different functionalities. One of the key attributes of Internet *standards* is that they focus on protocols and not on implementations. thus standards allow for services to be described in a common language web service description language(WSDL), to use a common messaging protocol to communicate (SOAP) and finally the integration of these web services and the creation of newer ones based on some features of existing ones (PBEL). The standards are a collection of specifications, rules, and guidelines formulated and accepted by the leading market participants and are independent of implementation details. Standards establish a base for commonality and enable wide acceptance through interoperability [6]. We are also using BPMN Business Process Modeling Notation to express the workflow of our scenario. Figure 3 illustrates the workflow diagram of our scenario. There are three participating web services: the LMS, YouTube, and RealOne downloader. The user submits the request for a video through the LMS. The figure3 shows the process to process coordination and communication, the composer application describes how a video request is executed by YouTube WS via the RealOne downloader. Several activities are executed by each of the participating WSs. YouTube WS has to check if RealOne downloader plugin is available (not included in the figure) if so it will submit the request to download it then submit the request to download the video selected by the user through the LMS. The RealOne downloader has to execute the request submitted to it by the YouTube WS. In the remainder of this paper we describe how the orchestrator coordinates the interactions between these different services and explain how it uses integrated policies to apply orchestration rules.

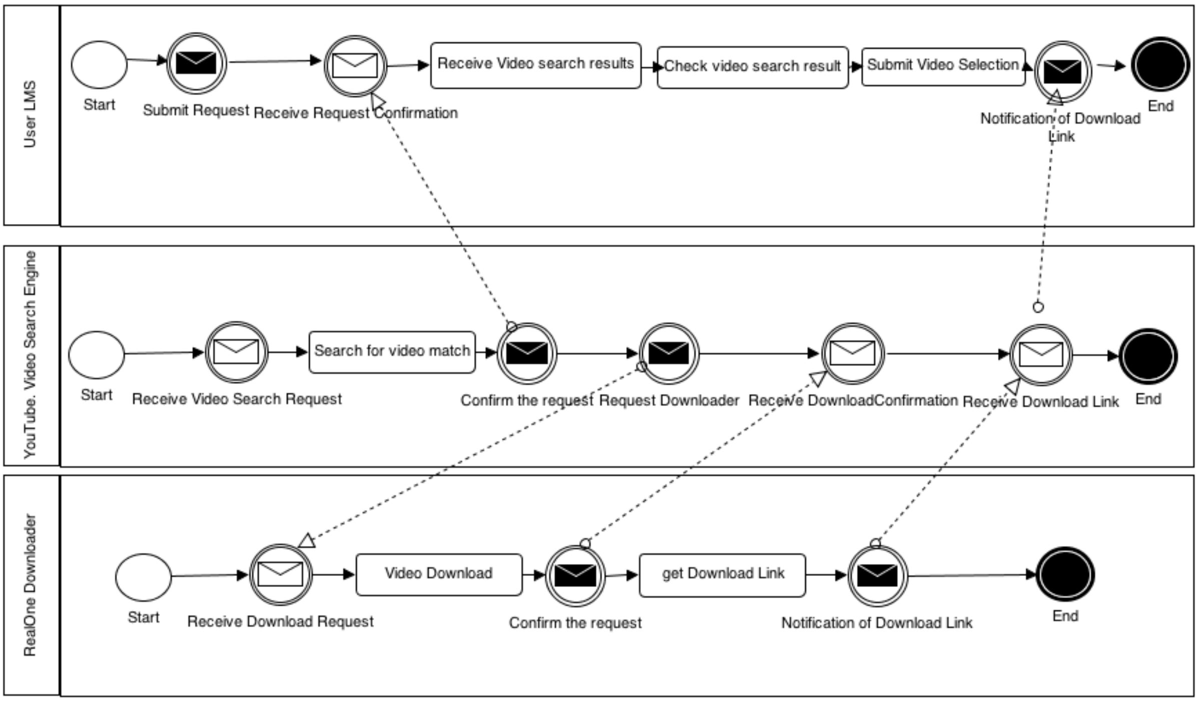


Figure3: YouTube download Request flow diagram

a) Finding a matching service:

Each participating service in the composition process is described using a WSDL (web service description language). WSDL is based on XML, some of the advantages of XML is that is retains the data structure in transit, it is also very flexible which solves application heterogeneity and middleware problems [6]. The WSDL file contains information about the service, how to access it and its input and output. WSDL is a specification defining how to describe web services in a common XML grammar. WSDL describes four main important pieces of data:

- Address location of the specified service
- Information about the interface to describe all service functions
- Information about the Data type for all message requests and message responses
- Information about binding specifications for the transport protocol used

We are using the service related data stored in the WSDL to search and locate a set of services using the beam stack algorithm. It is guaranteed that a match is found if one exists. The participating services are both language and platform independent. Services by different providers running in different platforms and written in different languages can interact and communicate easily using this specification. We are using the WSDL data to allow for a given service to be invoked by clients and to use all or some of its functionalities.

Figure 4 is a sample of the (LMS) Moodle Wsdl file. Moodle is one of the participating services in the service composition process. As stated previously, each participating service can contain several sub-services. For Youtube as a web service it contains: VideoSearch. unregistered users can watch most videos on the site; registered users have the ability to upload an unlimited number of videos; Flag – ability to indicate a video that has inappropriate content; Title - main title of the video; Tags – keywords specified by the person who has uploaded the video; Channels – relating to groupings of content; Related videos - determined by the title and tags, appear to the right of the video; Subscribe: registered users can subscribe to content feeds for a particular user or users[4]. We use VideoSearch which is a subservice as a participating service in the composition process.

Our Moodle WSDL defines the interface which comprises: the abstract interface description containing the supported operations, the operation parameters and their types; also it contains binding and implementation description containing a binding of the abstract description to a concrete transport protocol, message format, and network address. Moreover the WSDL will also include the policies specific to each service in order to enforce integrity and allow the service to be executed by different consumers without sacrificing security. We will talk about these policies in the next section 'service specific policies'

```
            <input>
                <soap:body use='encoded' namespace='urn:xmethods-delayed-quotes'
                           encodingStyle='http://schemas.xmlsoap.org/soap/encoding/
            </input>
            <output>
                <soap:body use='encoded' namespace='urn:xmethods-delayed-quotes'
                           encodingStyle='http://schemas.xmlsoap.org/soap/encoding/
            </output>
        </operation>
    </binding>

    <service name='userService'>
            <port name='userPort' binding='userBinding'>
                <soap:address location='http://jerome.moodle.com/Moodle_HEAD/moodle
            </port>
    </service>
</definitions>
```

Figure 4: the moodle wsdl file

i) Service Specific Policies:

One of the challenges in service composition is how to ensure that rules that are supposed to be applied to a specific service will still take place even if it is executed by outside sources. Especially, authentication and security rules. As a solution to this problem we express these rules as policies. These policies are included in the wsdl description of each service. This guarantees efficiency. An example of a service specific rule or policy is a restriction that a gradebook service has which doesn't allow modification of the gradebook values after the end of the term. This ensures that once a given term is completed, the grades in the gradebook are final and may not be modified. The orchestration platform gets the policy directly from the WSDL, given that the policy is defined in the WSDL. In our scenario, the videos that we search for on YouTube can be added to the LMS only for current courses, once the teaching term is finished all current courses are listed as past courses. Therefore the wsdl file for the LMS includes a policy that learning content can only be added for current courses (the status of a course is current)

The policy is directly loaded from the service description file, which is in our classpath, and will apply all policies in runtime, provided that this resides in the application server and doesn't have any conflicting dependencies in the project. Metro is one that allows such specifications. The service specific policies are sent via SOAP messages as a SOAP header.

b) Orchestration Algorithm

Web services seek to create interoperability among information systems. Web service orchestration enables the coordination of activities [7]. Orchestration is one approach to accomplish service composition in which there is a central server that controls and manages all the interactions between the participating services. The orchestration code is located in the central server. Once a user submits a request for a service that doesn't exist, the call to compose a service comes into play. The central server searches for a set of services that match at least part of the user's request. Based on the set of basic services, the central server then must order and manage the execution as well as the interactions between the participating services; next from a set of possible combinations that meet the user's request, the orchestrator must select the best one then execute it so that the user can use it as a single composed service. The execution of the composed service requires retrieval and exchange of data and code with each basic service provider. This execution can be centralized or distributed among multiple servers. There are several existing tools and standards related to service orchestration such as BPEL business process language. Although a BPEL program invokes services distributed over several servers, the orchestration of these services is typically under centralized control [8]. To perform a decentralized orchestration using BPEL, the orchestration code must be divided among multiple servers. When the BPEL program is partitioned we get multiple independent sub-programs. Those subprograms do not need centralized control to interact. There are also partitioning algorithms and programs like Zenflow and the one proposed in [8].

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 1, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

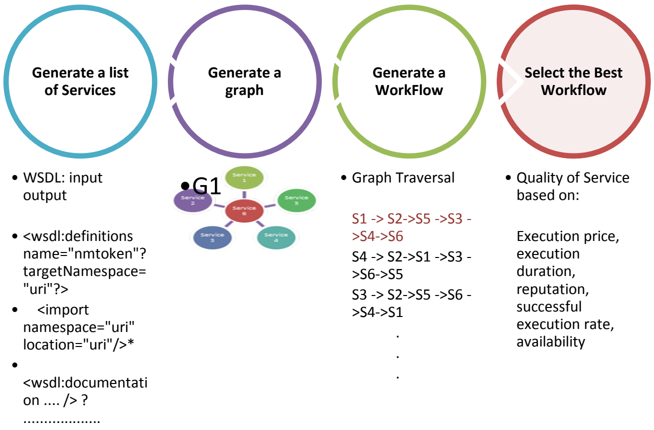


Figure 5: Service Composition Process using the proposed algorithm

Here we will illustrate how we get from step 2 of figure 5 to step 3 of the same figure. How we use the algorithm in figure6 to generate a set of possible workflows based on the results of our service selection (step1 of figure4). S is a set or participating services. G is the graph that contains all the services that match the request. Pn represents a set of processes generated. Each generated process represents a potential solution to the user's request. Each process is represented using a traversal of the graph where the vertexes are the policies that will control the flow of execution between services. In the orchestration process the built in policies within the WSDL of each of the participating services will be used to generate new policies to manage the flow of execution of the composed service. Those policies will be stored as vertexes of the graph.

```
FindAllProcesses()

        P1: S1->S2->S4
        P1: S1->S3->S4
        P1: S8->S4->S2->S5
        P1: S2->S4->S1

                                                        //a policy is generated when
                                                        // a transition is created
CreateProcess()
for each s in SS                                        //SS is the service set
      check against all other s in SS
        if match(s,s') then                             // s is a service x and s' is
                                                        // service y in the SS
                 MakeTransition(s,s')
                 Process.insert(P')                     // G is the graph (nodes are servic
                                                        // transition are policies
if FixedPoint(G) then
        print("reached fixed point")

MakeTransition(s,s')
        if s(I) = s'(o) then
                 insert(s,s')
        else
```

Figure 6: The orchestration algorithm

i) Communication between Services

Since we are supporting web service composition then communication between the participating services is done using the standard SOAP. SOAP stands for simple object access control. It is a standard to allow different services by different providers to communicate and exchange data. SOAP protocol is intended for exchanging structured information in a decentralized, distributed environment. It uses XML technologies to define an extensible messaging framework providing a message construct that can be exchanged over a variety of underlying protocols [10]. The framework is intended to be independent of any specific programming model and other implementation particular semantics. Figure7a and Figure7b show the soap binding used to create communication links between the web services.

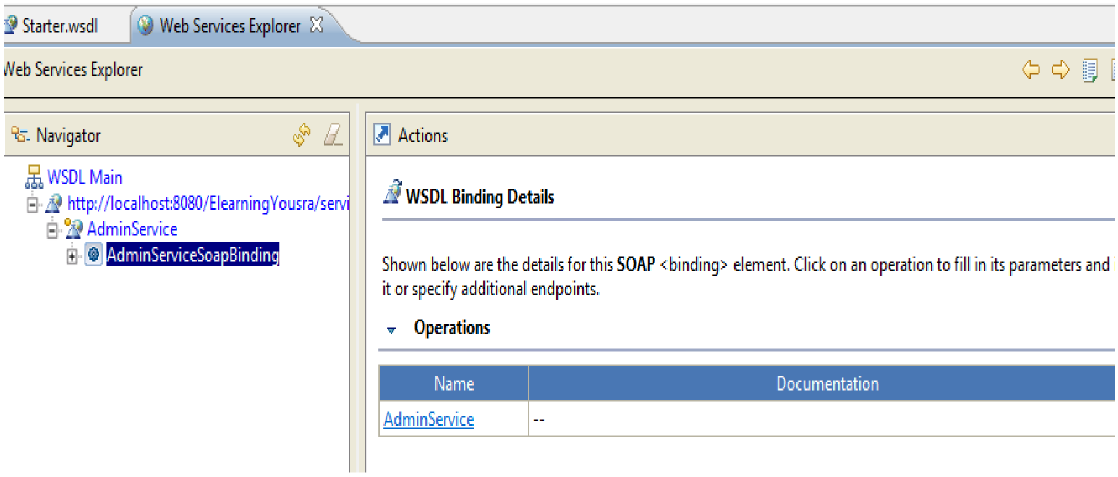


Figure7a: Example of messaging protocol is used to define binding between the basic services

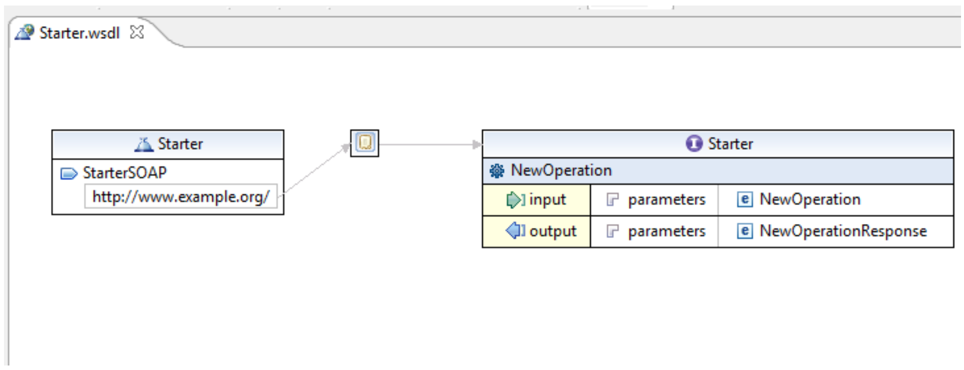


Figure 7b: Example of input/output definition within SOAP operations

ii) Orchestration specific policies:

These are rules and restrictions specific to the management of the orchestration process. Each policy will specify how a given service may be composed with another. Our services are described as nodes and the policies are the vertexes of that graph. Each policy may contain several alternatives.

Alternatives within a policy may differ significantly in terms of the behaviors they indicate. Conversely, alternatives within a policy may be very similar. In either case, the value or suitability of an alternative is generally a function of the semantics of assertions within the alternative [9] the orchestration algorithm allows generating policies that will handle the execution of the composed service. Those policies are derived from policies of basic participating services. Since WS-Policy offers a framework through which we can extend the description features already provided through the WSDL. WS-MetadataExchange protocol allows to dynamically exchange WS-Policy between two participating service endpoints. Using the WS-MetadataExchange protocol, service endpoints can exchange policies at run-time to bootstrap their interaction with information about the settings and protocols that apply [2]

iii) Validation of the Orchestration Algorithm:

The search for all possible matching services to the users request is performed using the beam stack search algorithm with backtracking; Beam search is a widely-used approximate search algorithm. By focusing its search effort on the most promising paths through a search space, beam search can find a solution within practical time and memory limits – even for problems with huge search spaces [11]. However the beamstack algorithm is not considered comprehensive because it won't necessarily find the optimal solution. This means that in order to guarantee that the optimal solution is to be found we opt to use beamstack search with backtracking. If there are no services that match the user's request; the process halts and notifies the user of the result. The orchestration algorithm generates a set of possible workflows based on all possible traversals of the graph. The algorithm is using graph traversal in which the orchestration policies are the vertices of the graph. This traversal will allow generating processes even when we have a set of services that match partially the user's request. In case there is a fault in the execution of one of the services, the system will continue with the next possible process in the list of generated processes (different traversals). The processes are all ordered by priority order based on one criteria (quality of composed service). Quality of composed service (process) is based on how close is the process to match the user's request. We should also mention that the use of policies allows more flexibility dealing with different participating WSs differently. It is customized for each of the WS as needed. With this kind of flexibility, WSs can then be designed to offer differentiated QoS (quality of service); eg: service precision, granularity, timeliness and scope, depending on the end customers. WS can also be differentiated based on technical quality and QoS details such as response time, performance bandwidth used, and reliability [2]

4 Conclusion and future work

In this paper we have presented an evaluation of a service orchestration algorithm for web service composition. Web Service composition has being addresses using different method and techniques one of those methods are rule based techniques. Although our method is rule based as well we are making use of WsPolicy as a web service rule standard. Also our algorithm allows the web service composition to be dynamic to a certain level by generating newer policies that will address the execution of the composed service. This resolves several issues including that only minimal changes are needed to the existing environment. Newer web services may be added dynamically if they are described using the same WS standard used. This will allow the system to expand continually and dynamically. We have tested our proposed algorithm using a simple E-learning scenario which is to search for learning content (a video from YouTube), download it then upload it to the LMS. Our method preserves the restrictions and conditions for execution of each participating web service. For future work a translator can be added to the current system in order to translate the user's request from natural language or any other service user language users choose to input their request in. The translator will then translate it into the service specification language used WSDL and therefore the system will be user-friendly.

References

[1] Y.Chtouki, H.Harroud, M.khalidi, S.Bennani ; 'Policy Based Orchestration in Service Composition' ;IJAEST-Vol-No-11-Issue-No-2; ISSN: 2230-7818; http://www.ijaest.iserp.org/ijaest-archieves-vol11.2.html ; 2011.

[2] M.Papazoglou ; 'Web Services & SOA : Principles and technology', Pearsoned ; 2012 ; Print

[3] The center for Learning & Performance Technologies; 'Top 100 Tools for Learning 2012'; http://c4lpt.co.uk/top-100-tools-2012/

[4] Duffy, P. "Engaging the YouTube Google-Eyed Generation: Strategies for Using Web 2.0 in Teaching and Learning." The Electronic Journal of e-Learning Volume 6 Issue 2, pp 119 - 130, available online at www.ejel.org

[5] Y.Chtouki, H.Harroud, M.Khalidi, S.Bennani; "The impact of YouTube Videos on the Student's Learning"; Proceeding of ITHET 2012;Turkey; 2012.

[6] W.Roshen; 'SOA-Based Enterprise Integration';McGraw-Hill;2009;Print

[7] Janssen, Marijn, Kuk, George,' Comparing Coordination Arrangements Enabled by Web Services and Web Service Orchestration Technology'

[8] M.Nanda, S.Chandra, V.Sarkar; 'Decentralized Execution of Composite Web Services';

OOPSLA '04 Proceedings of the 19th annual ACM SIGPLAN conference on Object-oriented programming, systems, languages, and applications; Pages 170-187; USA; 2004.

[9] S.Bajaj et al; 'Web Services Policy 1.2 - Framework (WS-Policy)';W3C Member Submission; April 2006

[10] M. Gudgin, et al, 'SOAP Version 1.2 Part 1: Messaging Framework'; June 2003.

[11] R. Zhou, E. Hansen, Beam-stack search: Integrating backtracking with beam search, in: Proceedings of the 15th International Conference on Automated Planning and Scheduling (ICAPS-05), Monterey, CA, 2005, pp. 90–98.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 1, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

Modeling Methodology for NoC: MM4NoC

Nassim Sefrioui[1], Aissam Berrahou2 , Pr. Diouri ouafae[3] and Mohsine Eleuldj[4]

**[123] University Mohamed V, Mohammadia School of Engineers,
Rabat , Morocco
[4] National School of Applied Sciences, PO Box 669,
Oujda 60000, Morocco**

Abstract

The current technology allows the integration on a single chip of complex systems "SoC" which are composed of pre-designed blocks "IPs" that can be interconnected by a network on chip "NoCs. Generally, the IPs are validated by various techniques "simulation, test, formal verification" and the main problem remains the validation of communications infrastructure in order to compare the performance - latency, throughput, power consumption, area occupied.

Based on the work done in [1] that focuses on the formal verification of networks on chip using an automatic proof tool, this article presents a draft introduction to behavioral modeling and simulation of a communication network on chip.

Keywords: *communication network on chip, Multiprocessing, on-chip, N-dsp, N-core system.*

1. Introduction

This article provides an introduction to modeling and behavioural simulation of a communication network on chip "NoC - Network on Chip" [3].

The rapidly changing and increasingly complex systems on a chip "SoC - System on Chip" to SoC multi-processor "MPSoC - Multiprocessors SoC" interconnection communication modules or cores constituting these systems has undergoes a topological and structural change.

The latter meets the constraints of performance and cost related to the complexity and the increasing modules or interconnected cores. Currently this trend is moving towards the integration of on-chip communication network, implementing the transmission of packet data to the interconnected network nodes and corresponding to the "processor modules, memory controllers, connected devices,... "

The transmission of data is done through routers constituting the network and implementing rules switching and routing packets across the network [2].

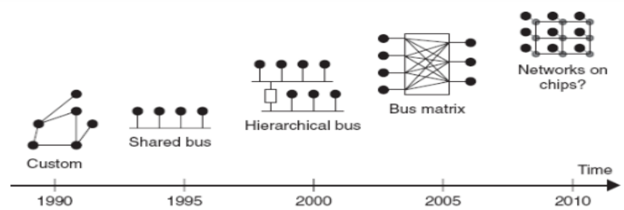


Fig. 1 Evolution in the interconnected SoC

2. Description environment of a NoC

2.1 Description NoC

Regarding technology, the term most commonly used is that of network-on-chip " NoC abbreviated as" derived from system-on-chip "SoC". While applications such as network on silicon or on chip network are rarely used.

Research on network-on-chip integrates various fields of development.

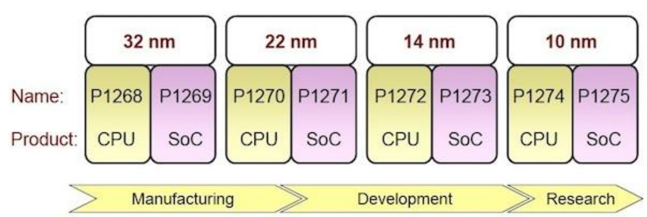


Fig. 2 NoC: from research to industrial use (By [7]).

Currently, all aspects of NoC have not yet been addressed and there is no industrial solution with a level of maturity comparable to the bus. NoCs are still at the research **Fig.2**.

One of the major advantages of on-chip networks is to provide a reliable and efficient interconnection structure to rapidly develop new routes from reusable building blocks, previously developed and validated, which provide a system for exchanging data between different processing resources: processors, memory, DSP "Digital Signal Processor" or cores...

2.2 Topology network interconnection

There are a wide variety of network topologies [3], whether direct or indirect, regular or irregular. Obviously, the list of topologies that follows is not exhaustive but represents those most often used in NoCs.

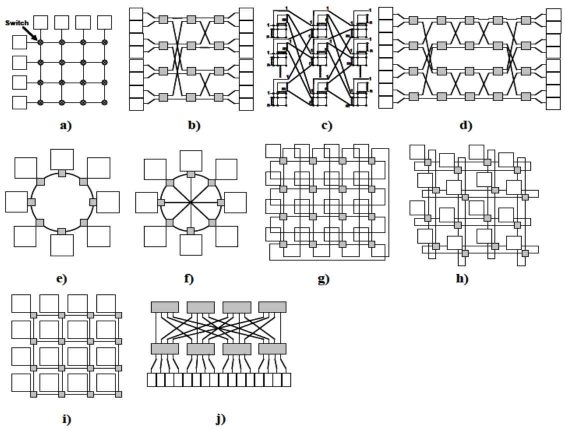
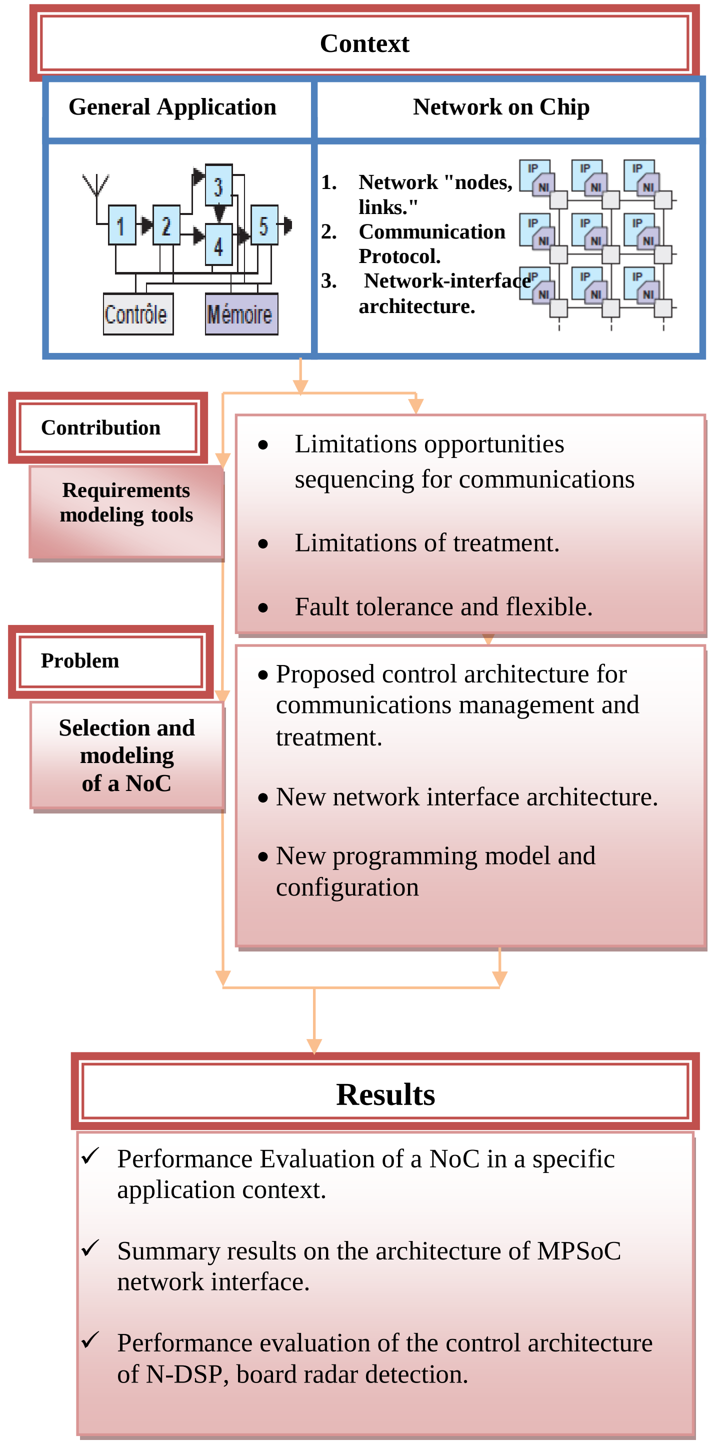


Fig. 3 Topologies often used in Nocs: (a) crossbar (b) Buttery (c)Clos (d)Benes (e) ring (f) chordal ring (g) torus (h) folded torus (i) 2D mesh (j) fat tree. **[3]**

Network interfaces and routers can be configured to optimize communication paths of data flows exchanged through the NoC. There are several examples of network-on-chip: STNoc , SPIN , Xpipes , Hermes, FAUST Æthereal

3. Diagram of Research

This part of research is the study and modeling of a network on chip. architecture in the context of the implementation.



MODELING THE NOC, It models the behaviour of the communication network of independent processing units that will use this network. Several approaches are possible:

3.1 Use modeling environment high-level

The goal is to quickly develop a working model without the level of detail required for a physical implementation. These environments provide a set of primitives to build a system and simulate it. Include work done with Ptolemy [10] as SSMK01[12] and ZhMa02[13] or SDL [11] "specification and description language" for [6] and [5]. Overall, these environments are less used because they are not known in the middle of the design.

Most designs are planned and written in VHDL designed to be implemented directly on a chip usually a CPLD or FPGA.

3.2 Use a tool for modeling network

A solution to simulate a network on chip architecture is to take standard tools. Several network simulators are available and can be adapted to the problems of NoC. E.g. ModelSim [8], OPNET [9].

3.3 Develop a specific model

The approach most often chosen to simulate the operation of a network on chip is to develop a specific model of the architecture is designed from description languages or programming standards in order to have a model specific cycle or even the bit level to create an interconnection network onboard FPGA interconnects N-DSP of N-core in a MPSoC platform.

4. Evaluating the performance of NoC

4.1 General methodology

We distinguish four distinct phases: modeling (representing the behaviour of the system), programming, testing and interpretation of results (with shares).

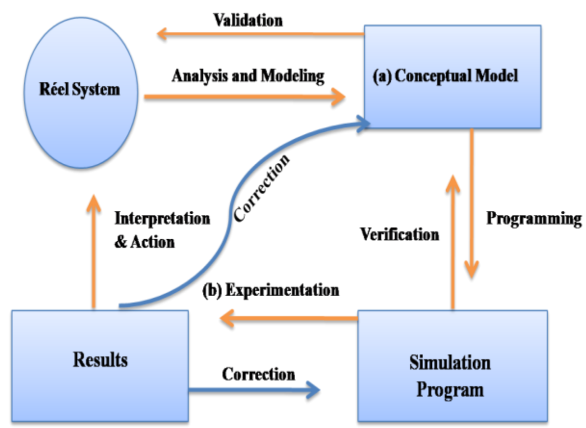


Fig. 4 Methodology adapted

- (a) **Conceptual Model:** The model is only an approximation of the system; it is conditioned by the objective of the study.
- (b) **Experimentation:** This build theories or assumptions that take into account the observed behaviour.

4.2 Implementation and performance analysis

CASE STUDY: achieve an interconnection network between four DSP chips Tms320c6474 three cores each and as much memory [4] different testing capabilities offered by this device to make. The platform has a CPLD future location of a NOC. [4bis]

At the beginning the study is to carry forward two DSP cores of two different but on the same map simulation, the result is a transfer pack data 256 bytes, via "The C6474 Evaluation Module" software.

5. Conclusions

This study modeling, simulation and communication network on chip architecture assessment to raise the fundamental role on the performance of interconnected systems expected in a MPSoC platform in its design.

The object is on the one hand the development and design of microelectronic interconnects, constituting the key performance of a multiprocessor platform, secondly, to show the value of using modeling tools allowing the simulation in order to test and validate an environmental architecture in its operating environment.

Indeed, the development of systems that are becoming more complex with the integration of both hardware and software parts, the technical co design.

The idea to connect and communicate several processor cores in a chip, which is a cross-disciplinary work

"Networks, Nanotechnology, Embedded Systems and Digital Communications " is:

- A reliable solution for modeling on-chip communication networks.
- A prototyping methodology based reconfigurable circuits (ie FPGA / CPLD) for rapid validation of systems design.
- Build on the results across multiple applications.

References

[1] Behavioral modeling and C-VHDL co-simulation of Network on Chip on FPGA for Education C. Killian, C. Tanougast, M. Monteiro, C. Diou, A. Dandache LICM, University Paul Verlaine of Metz Metz, France . May 17-19, 2010, Karlsruhe, Germany

[2] THESE Amr HELMY le 30 April 2010 "Implementation techniques for the automatic demonstration Formal verification of NoCs page 6

[3] G. De Micheli et L. Benini, « Networks on Chips, Technology and tools », Morgan Kaufmann publishers, 2006.

[4] N.sefrioui, M.Eleuldj Studies multiprocessor platforms at base DSP/Tms320c6474, Communication intercores, JDSIR '2010 Mohammadia School of Engineers.

[4bis] N.sefrioui, M.Eleuldj Communication intercores with network on chip Implemented FPGA device. JDSIR '2012 Mohammadia School of Engineers.

[5] D. Andreasson et S. Kumar – « On improving best-effort throughput by better utilization of guaranteed throughput channels in an on-chip communication system » proc 22nd IEEE norchip conf nov 2004, P 265-268

[6] R. Holsmark, M. Hogberg et S. Kumar – « Modeling and evaluation of a network on chip architecture using sdl », Proc. 11th SDL Forum, July 2003

[7] International Technology Roadmap for Semiconductors.2005Edition. http://www.itrs.net/Common/2005ITRS/Home2005.htm

[8] High Performance and Capacity Mixed HDL Simulation – ModelSim http://model.com/

[9] OPNET Network Performance Management solutions OPNET http://www.opnet.com/

[10] The Ptolemy project studies modeling, simulation, and design of concurrent, real-time, embedded systems. http://ptolemy.eecs.berkeley.edu/

[11] The SDL Forum Society http://www.sdl-forum.org/

[12] M. Sgroi, M. Sheets, A. Mihal, K. Keutzer, S. Malik, J. Rabaey et A. Sangiovanni-Vincentelli « Addressing the system-on-a-chip interconnect woes through communication based design »

[13] X. Zhu et S. Malik – « A hierarchical modeling framework for on-chip communication architectures »

First Author. Engineer on Ministry of Higher Education. Received the DESA I.T.M in computer science, Telecom and Multimedia from University Mohammed V Agdal , EMI RABAT Morocco. He is currently a researcher on Mohammadia School of Engineers, EMI Rabat Morocco. Her research focuses on interconnection networks on chip and parallel computer architectures, design methodologies for nanoscale systems on-chip, with a special interest on network-on-chip communication architectures.

Second Author. Administrator on MEN. Received the DESA in computer science from University Mohammed V Agdal, EMI RABAT Morocco. He is currently a researcher on Mohammadia School of Engineers, EMI Rabat Morocco. His research interests are in the fields of embedded systems and electronic design automation. In particular, her research focuses on the co design, SW / HW.

Third Author. PES received the D.E.S in computer science from the University Paris XI FRENCH in 1985.She is currently Professor of Mohammadia School of Engineers, EMI RABAT Morocco. Her research interests include communication architectures.

Fourth Author. Director ENSAO received the Ph.D. degree in computer science from the University Montreal CANADA in 1989. He is currently Professor and Director of National School of Applied Sciences - Oujda, Morocco. His current research focuses on the development of design methodologies and software tools for the design of systems on chip, the on chip communication.

Importance of Software Documentation

Noela Jemutai Kipyegen[1] and William P. K. Korir[2]

**[1]Department of Computer Science, Egerton University
Njoro, Kenya**

**[2]Department of Computer Science, Egerton University
Njoro, Kenya**

Abstract

Software Documentation is a critical activity in software engineering. Documentation improves on the quality of a software product. It also plays significant roles in software development environment and system maintenance. Several software development best practices are ignored. This paper looks at the importance of software documentation, creation, applications, tools and standards.

Keywords: *Software Documentation, Importance, Role, Applications, tools and standards*

1. Introduction

Many factors contribute to the success of a software project; documentation included. *Software documentation is an artifact whose purpose is to communicate information about the software system to which it belongs* [1]. Parnas [2] defines *a document as a written description that has an official status or authority and may be used as evidence. In development, a document is usually considered binding, i.e. it restricts what may be created. If deviation is needed, revisions of the document must be approved by the responsible authority*.

Systematic approaches to documentation increase the level of confidence of the end deliverable as well as enhance and ensure product's success through its usability, marketability and ease of support [3]. "The dominant factor between a successful project and an unsuccessful project reduces to the effective dissemination of key information and successful software projects become successful because they give the right level of attention to clearly communicating the key concepts and requirements" [4].

Capri defines a successful documentation as one that makes information easily accessible, provides a limited number of user entry points, helps new users learn quickly, simplifies the product and helps cut support costs.

Poor documentation is the cause of many errors and reduces efficiency in every phase of a software product's development and use [2]. Documentation is an activity that needs to commence early in development and continue throughout the development lifecycle. It acts as a tool for planning and decision making.

2. Motivation

After assessing students' projects for a period of time, we realized majority of the students are neither enthusiastic nor motivated in the area of documentation. Most of them prefer only one phase of software development which is, coding. From this issue, we developed a desire to deeply understand software documentation: applications, benefits, creation and role in software development environment. *Does it contribute to the success of a project?* This led us to explore the existing documentation practices in software engineering.

3. Document Creation

Capri [3] describes in Figure 1, eight processes (analysis, design, development, validation, production, manufacturing, delivery and customer satisfaction) that guide in document creation. *Document preparation is the process of creating a document and formatting it for publication* [5]. Other researchers [6] gave seven rules for sound documentation. These rules include; 1. Documentation should be written from the point of view of the reader, not the writer, 2. Avoid repetition, 3. Avoid unintentional ambiguity, 4. Use a standard organization, 5. Record rationale, 6. Keep it current and, 7. Review documentation for fitness of purpose.

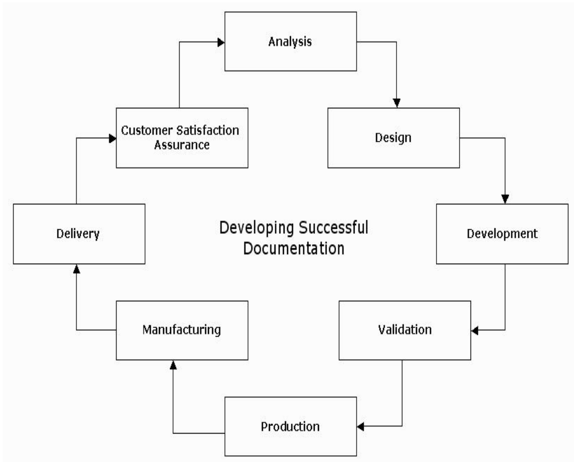


Fig. 1 Eight Phases to successful documentation (source: [3])

In the first Phase, analysis, possible audience that could potentially need documentation for the product and the tasks they will perform on the software are identified. Next phase, design, involve taking all the documentation items identified during the Analysis Phase and contents for each are designed/planned. The third phase entails creation of the actual document to be delivered. Validation, phase five, entail testing the documentation to ensure it meets its performance objectives, and the needs of its target audiences. The purpose of phase six is to produce high-quality finished goods (paper, videotape, audio, CD, online, etc.) to meet demand. In the seventh phase, final product (software and documentation) is then delivered to the customer. The last phase is customer satisfaction. In this phase, the document is improved based on customer's needs [3]. Sommerville [5], described document preparation process in three stages namely document creation, polishing and production as shown in Figure 2.

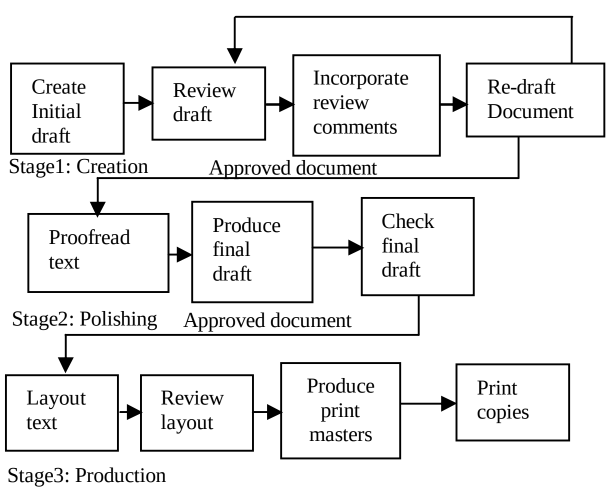


Fig. 2 Stages of document preparation (Source: [5])

Document creation involves initial input of the information in the document. The second stage which is document polishing entails improving the writing and presentation of the document to make it more understandable and readable. The last stage is document production which defines the process of preparing the document for professional printing [5].
The two models, (Capri's and Sommerville's) define software document creation as a process which involve continual understanding, review, and modification throughout the development lifecycle [7] as shown in Figure 3 below.

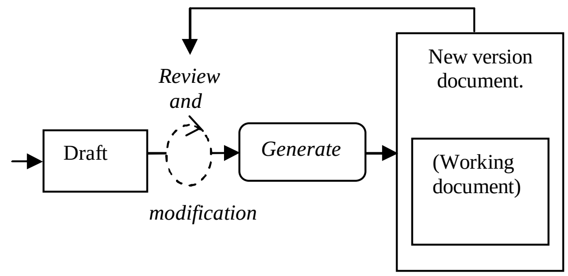


Fig 3 Document creation

Apart from being documentation cost and benefit model [8], Bo sun's model (Figure 4 below) describe in-depth the processes involved in software documentation. The model gives clear illustration of various actors (e.g Requirements Engineer, Business analyst, document reviewers, software developers and maintainer of the system) involved in software development, and uses of the documents. Bo Sun's model is comprehensive and can be efficiently employed in the description, creation, applications, as well as understanding the costs and benefits of documentation.

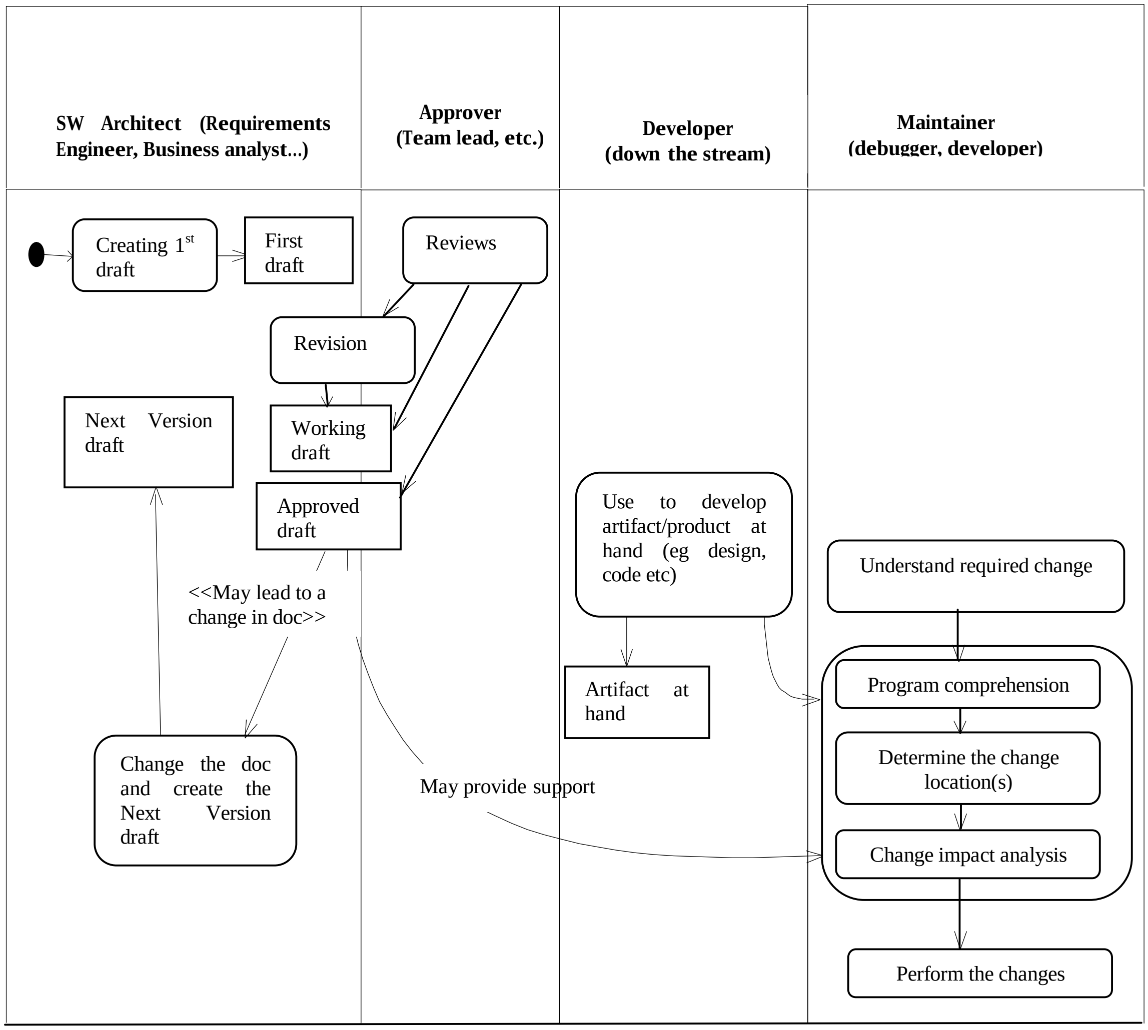


Fig. 4 Documentation cost and benefit model (Source: [8])

4. Application and benefits of documentation

The role of documentation in a software engineering environment is to communicate information to its audience and instill knowledge of the system it describes [1]. As seen in Bo Sun's model, system documentation plays another role namely software maintenance.

Parnas [2] identified several benefits of documentation. These include: easier reuse of old designs, better communication about requirements, more useful design reviews, easier integration of separately written modules, more effective code inspection, more effective testing, and more efficient corrections and improvements.

Researchers and practitioners also have looked at the uses of software documentation and just to compile a few, [9] notes that documentation helps at software development, keeps software-quality at high levels and makes it easy to transfer projects. Documentation can also be used for [10] learning a software system, testing a software system, working with a new software system, solving problems when other developers are unavailable to answer questions, looking for big-picture information about a software system, maintaining a software system, answering questions about a system for management or customers, looking for in-depth information about a software system, working with an established software system. Akin-Laguda [11] lists other uses which include; facilitates effective communication regarding the system between the technical and the non technical users, training new users, solve problems like trouble shooting, evaluation process, and quantify the financial ramifications/footprint of the system.

All the above mentioned uses of documentation, can be simplified and summarized as [5] puts; 1. Documentation is used as a communication medium between members of the development team and probably the clients, 2. Used for maintenance, 3. Provide information for management to help them plan, budget and schedule the software development process and, 4. Tell users how to use and administer the system.

5. Types of Documentation

Sommerville describes two main categories of software documentations; process and product documents. Process documentations are used to manage the development process for example planning, scheduling and cost tracking, standards among others. Product documentations describe the main deliverable (software product) and some of the documents in this category form part of deliverables. These include; Requirements Specification, Design documents, Commented Source Code, Test Plans including test cases, Validation and Verification plan and results, List of Known Bugs and user manual [5].

6. Tools for documentation

Forward and Lethbridge [12] in their survey found the following documentation tools more helpful; MS Word (and other word processors), Javadoc and similar tools (Doxygen, Doc++), Text Editors, and Rational Rose.

Doxygen is a documentation system for ***C++***, C, ***Java***, Objective C, Fortran, VHDL, PHP, C#. Doxygen generates; on-line documents in HTML, off-line manual in latex and output in RTF (MS-Word), PostScript, hyperlinked PDF, compressed HTML, and Unix man pages. It extracts documentation directly from the sources, which makes it much easier to keep the documentation consistent with the source code. Doxygen can be configured to extract the code structure from undocumented source files. This is very useful to quickly locate elements in large source distributions. Doxygen include; dependency graphs, inheritance diagrams, and collaboration diagrams, which are all generated automatically. These help to visualize the relations between the various elements. Doxygen can be used for creating normal documentation. Another important aspect of Doxygen is that, it is highly portable [13].

A word processor is a program that is used to produce, edit and format text. “Word processing systems are screen based. This means that the image of the document on the user’s terminal is, more or less, the same as the final form of the printed document. Layout can be improved before printing the document”[5].

LATEX is a TEX macro package that simplifies the use of TEX. It is portable and come with its set of fonts. Most LATEX commands are “high-level” and specify the logical structure of a document. This simplifies the authors work by taking care of document layout details. LATEX provides several standard document classes from which to choose. The document class determines how the document will be formatted. Other benefits of LATEX include; its flexibility, gives the user complete control, handles big, complex documents with ease, and never crashes or corrupts users’ files [14].

7. Software Documentation standards

Standardized documentation can be defined as documents having a consistent appearance, structure and quality. This means should be easier to read, understand and usable [5], [15]. Standards act as a basis for document quality assurance.

“Using a standard means that documentation producers and customers have a consistent accepted reference for the format and content that they will find in the documentation. For example, what documentation must be printed? What does it mean to say that the documentation is “complete”? Does it have to include every function and screen shot?” [15].

Reilly discusses various ISO Software documentation standards which include;

ISO/IEC/IEEE 26514:2008, *Systems and software engineering-Requirements for designers and developers of user documentation*. This standard details both process and product standards.

ISO/IEC/IEEE 26513:2009, *Software and systems engineering-Requirements for testers and reviewers of user documentation*. This standard covers the activities and responsibilities for planning and conducting documentation reviews and managing the results of the review. It also addresses how to plan, measure, and conduct usability tests for documentation, along with tests of accessibility and of localized or customized versions.

ISO/IEC/IEEE 26512, *Software and systems engineering-Requirements for acquirers and suppliers of user documentation.* It lays out the processes for acquiring user documentation services and for monitoring and managing contractors.

IEEE Std 1063-2001, *IEEE Standard for Software User Documentation.* This standard is a revision of IEEE std 1987. It provides minimum requirements for the structure, information content, and format of user documentation, including both printed and electronic documents used in the work environment by users of systems containing software. This standard is limited to the software documentation product and does not include the processes of developing or managing software user documentation; it applies to printed user manuals, online help, and user reference documentation [16].

8. Conclusion

Software documentation is an activity of creating documents which are used in software development environment to communicate functions, operations and events to various stakeholders, for example software Requirements Engineers, Reviewers, Developers, operators, Maintainers of the system among others. Documents describe the product at all levels of development including the finished product. The documents also act as evidence of all the procedures and activities involved in software development therefore, documents need to be up-to date, complete, consistent and usable. To achieve consistency, systematic ways of document creation should be employed.

References

[1] A. Forward, “Software Documentation – Building and Maintaining Artefacts of Communication”. MS thesis. Institute for Computer Science, Ottawa-Carleton. Canada. 2002.

[2] Parnas, D.L.: Precise documentation: The key to better software. In: Nanz, S. (ed.) The Future of Software Engineering, 2011, pp. 125–148. Springer, Heidelberg

[3] S. Capri, Developing Successful Software Documentation, 2006. [Accesed: 14th August,2013], Available online:<http://www.softwareceo.com/downloads/files/sceo/white papers/DevelopingSuccessfulSoftwareDocumentation.pdf>

[4] Elowe, The role of documentation in software development. 2006. [Accessed: 9th July,2013]. Available online:< https://blogs.oracle.com/elowe/entry/the_role_of_documentation_in>

[5] I. Sommervile, "Software Documentation", 2001.Available from World Wide Web: <http://www.literateprogramming.com/documentation.pdf>.

[6] F. Bachmann, L. Bass, J. Carriere, P. Clements, D. Garlan, J. Ivers, R. Nord, R. Little, Software Architecture Documentation in Practice: Documenting Architectural Layers. CMU/SEI-2000-SR-004. Carnegie. Mellon University. 2000.

[7] F. Shull, “Developing Techniques for Using Software Documents: A Series of Empirical Studies”. PhD thesis, Computer Science Department, University of Maryland, USA. 1998.

[8] Bo Sun, “A Methodology for Analyzing Cost and Cost-Drivers of Technical Software Documentation”. MS thesis, University of Calgary. Alberta. 2012.

[9] A. Berger Guidelines: Technical Documentation Standard for Software development. Specific Group software solutions. v 1.3. 2010.

[10] T. C.Lethbridge, J.Singer, A.Forward, How Software Engineers Use Documentation: The State of the Practice, IEEE Software, v.20 n.6, 2003.p.35-39, [doi>10.1109/MS.2003.1241364]

[11] F. Akin-Lagida, Documentation in programming. NASSCOM. 2013. [Accessed online: 18th July, 2013]. Available:<http://www.slideshare.net/itniketan/importance-of-documentation>

[12] A. Forward and T.C. Lethbridge, "The Relevance of Software Documentation, Tools and Technologies: A Survey," *Proc. ACM Symp. Documentation Eng.*, ACM Press, 2002, pp. 26-33.

[13] Doxygen [Accessed: 16th July, 2013],Available:<http://www.stack.nl/~dimitri/doxygen/>

[14] Academic and Research Computing, Text Formatting with LATEX. A Tutorial. RPI, 2007. [Accessed on 14th August, 2013]. Available online:<www.unc.edu/depts/econ/egsa/LaTeX.pdf>

[15] A. Reilly, Audience-Oriented Standards for Software Documentation from ISO. Intercom; Vol. 58 Issue 3, 2011, p14-17

[16] IEEE, *Standard for Software User Documentation*, IEEEStd1063-2001. New York: Institute of Electrical and Electronics Engineers. 2001.

Noela Jemutai Kipyegen holds Master of Science in Software Engineering, awarded by Jomo Kenyatta University of Agriculture and Technology, Kenya in 2010. Bachelor of Science in Computer Science, from Egerton University, Kenya, awarded in 2006. Currently, she is an Assistant Lecturer and Researcher in the Department of Computer Science, Egerton University, Kenya. Research interests include Software Engineering Project Management, ICT for Development, looking at Human Computer Interaction (HCI). Publication: Kipyegen Noela J., Waweru Mwangi, Stephen Kimani,

Risk Management Adoption Framework for Software Projects: A Case Study for Kenyan Software Project Managers and Developers. International Journal for Computer Science Issues (IJCSI) Vol 9, Issue 3, May 2012.

William Paul Kiplangat Korir is a senior Lecturer at Egerton University. He holds MSc. in Computer Science from University of Regina, Canada, BSc. (Hons), University of Nairobi, Kenya. He has published several papers found in his profile, <http://www.egerton.ac.ke/index.php/Computer-Science/kiplangat-korir-william-pau.html>.

A Symbol Based Graphical Schema Resistant to Peeping Attack

T.Srinivasa Ravi Kiran[1], Dr.K.V.Samabasiva Rao[2], Dr.M.Kameswara Rao[3],A.Srisaila[4]

**[1] Lecturer, Department of Computer Science
P.G.Centre, P.B.Siddhartha College of Arts & Science,
Vijayawada, Andhra Pradesh, 520010, India**

[2] Principal, M.V.R College of Engineering, Paritala, 521180, Andhra Pradesh, India

[3] Associate Professor, Department Of Electronics and Computer Engineering, KL Univeristy,Vaddeswaram, 520 002,India

[4] Assistant Professor, Department of Information Technology, V.R.Siddhartha Engineering College,Kanur,Vijayawada,520006Andhra Pradesh,India

Abstract

Alphanumeric passwords are the most commonly used way of authenticating users in computer systems. One of the disadvantages of alphanumeric passwords is that they are hard to remember. Image passwords have been proposed to aim to make passwords more memorable and easier for users to use and, for this reason, it is more secure. Furthermore, most of the existing graphical password schemes are susceptible to spyware and shoulder surfing. This paper proposes a new graphical password authentication scheme resistant to peeping attack. In this schema We also try to answer two important questions: “Are graphical passwords as secure as text-based passwords?”; “What are the major design and implementation issues for graphical passwords?”. This survey will be useful for information security researchers and practitioners who are interested in finding an alternative to text-based authentication. methods. An analysis of security and usability aspects of the proposed scheme is presented.

Keywords: Graphical password; Authentication, Peeping attack, Security, Spyware

I. INTRODUCTION

Traditionally, alphanumeric passwords have been used for authentication, but they are known to have security and usability problems. The main problem with the alphanumeric passwords is that once a password has been chosen and learned the user must be able to recall it to log in. But, people regularly forget their passwords. If a password is not frequently used it will be even more susceptible to forgetting. To resist bruteforce search and dictionary attacks, users are required to use long and random passwords. Unfortunately, such passwords are hard to remember[1]. Furthermore, textual password is vulnerable to shoulder-surfing, hidden-camera and spyware attacks. Graphical password schemes have been proposed as a possible alternative to text-based schemes, motivated partially by the fact that humans can remember pictures better than text [2]. If the number of possible pictures is sufficiently large, the possible password space of a graphical password scheme may exceed that of text based schemes and thus presumably offer better resistance to dictionary attacks. Because of these advantages, there is a growing interest in graphical password. In addition, the possible password space of a graphical password scheme may exceed that of text based schemes and thus most probably offer higher level of security. It is also difficult to devise automated attacks for graphical passwords. As a result, graphical password schemes provide a way of making more human-friendly passwords while increasing the level of security. Due to these advantages, there is a growing interest in graphical password. However, existing graphical passwords are far from perfect. Typically, system requirements and cost of communication for graphical passwords are significantly higher than text-based passwords. In addition, few graphical systems support keyboard inputs. More importantly, most current graphical passwords are more vulnerable to shoulder-surfing attacks than textual passwords. In this paper we propose a textual graphical password system resistant to peeping attack. User study is conducted to explore the usability of the proposed scheme in terms of accuracy, efficiency and memorize ability. The rest of this paper is organized as follows. Section 2 briefly discusses related works on Graphical password schemes. Section 3 presents our proposed scheme. Section 4 examines usability issues and Section 5 deals with conclusion and future directions.

II. Related Work

In general, the graphical password techniques can be classified into in to three main categories: Locimetric, Drawmetric and Cognometric [6]. Locimetric authentication is an approach that exploits memorization and cued recall. This approach requires the user to use a background image to locate a series of predefined points. In 1996, Blonder [7] patented an innovative graphical authentication scheme called Graphical Password which is based on cued recall. In his design, the system first picks an image with many simple distinguishable locations, and these locations are stored on the system database. Wiedenbeck et al. [8] proposed and implemented an improved graphical authentication system called PassPoints.PassPoints is based on Blonder's idea of representing the password by multiple clicks on a single image. Drawmetric authentication is an approach that requires the user to draw a simple outline of the password during registration, and the user must redraw the similar drawing to be authenticated. Jermyn et al. [9] proposed and implemented a graphical authentication technique called Draw-a-Secret (DAS), which is primarily intended for devices with stylus input, such as Personal DigitalAssistants (PDAs). The main idea of DAS is that the user draws secret drawing (password) on a grid and the system verifies the drawing by checking the directions and the positions of the drawn strokes on the grid. Cognometric authentication is an approach that requires the user to identify a series of recognized images amongst a larger set of decoy images. Real User Corporation [10] developed a graphical authentication technique called Passfaces. The motivation behind Passfaces is based on humans' proficient ability to recognize human faces. Dhamija [11] has mentioned a major problem in authentication that users tend to have difficulties memorizing secure passwords.To overcome such problem, Dhamija et al. [12] suggested a solution called Déjà Vu, which improves the security of the system by replacing the precise recall of a text password with the recognition of seen images. Graphical authentication suffers a major drawback from Shouldersurfing. Shoulder-surfing refers to someone observing the user's action as the user enters a password. With graphical authentication, the user must select the recognized pictures from the displayed screen during login. Due to this, the user's action can be monitored by the attacker or it can be captured using recording devices such as camera. Wiedenbeck et al. [13] suggested a graphical password scheme for user authentication on computer called Convex Hull Click (CHC); and it was design to prevent shoulder surfing. Pierce et al. [14,15] proposed a technique that improves password security without additional hardware. Their technique exploits the ability of people being proficient to recognize visual information. Their proposed graphical authentication is called Authentigraph. The system first presents an image of randomly allocated artifacts on screen. Users are required to locate and select the recognized artifacts in sequence as their graphical password. De Angeli et al. [16,17] proposed a graphical authentication concept called Visual Identification Protocol (VIP) that aimed at improving user authentication in self-service technology. The notion of VIP is to replace the precise recalling of numerical code with the recognition of previously seen images for authentication. De Angeli et al. have suggested three prototypes of VIP, named VIP1, VIP2, and VIP3. Jansen et al. [18-20] proposed a visual login technique called Picture Password, which is designed for mobile devices with stylus input such as Personal Digital Assistants (PDAs).Hinds el al. [21] proposed a graphical password system called ToonPasswords. It requires users to select individual images from screens, which is similar to Passfaces [22] and Déjà Vu [23]. However, most of the current graphical password schemes do not have a balance between usability and security aspects. For example, if the system is too simple then the system may not be secure enough. If the algorithm is too complex then the system may not be user friendly, e.g.: difficult to learn and takes too long to log in.

III. Proposed Scheme

In this proposed scheme we use 5 x 5 grid formed using 25 blocks. Each block consists of a symbol. The symbol contains a set of four characters. The characters may numbers between 0 to 9, A to Z(Uppercase), a to z (Lowercase), Spaces and some special characters totally 95 character and 5 blank spaces are represented as shown in the Fig1. Passwords are input by typing or by mouse clicks.

Rules to write straight line on the interface

Fig 1 Proposed Schema

Rule 1: If user draws a line between two adjacent blocks, the password contains at least one character from each set of four characters depicted on the symbol of each block.

E.g. User draws a line between the top leftmost two blocks (row 1, column 1 and row 2, column 2), then the user would click on at least one character from each of the two sets [7,w,|,Q] and [8,!,P,blank-space] depicted on the blocks.

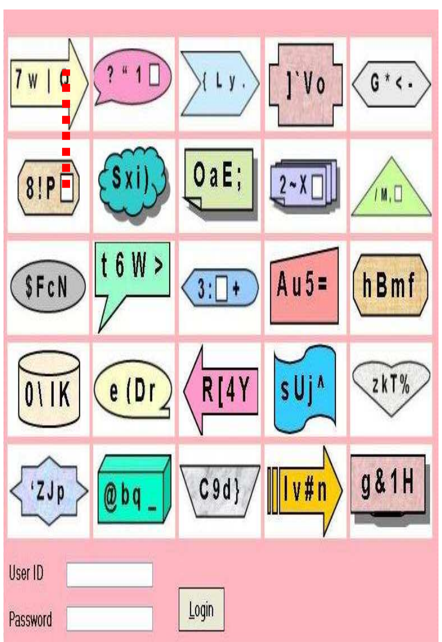


Fig 2 Line between adjacent blocks

Rule2: If user draws a line between two non-adjacent blocks, the character sets to be considered are taken from the symbol of the individual blocks in between if the line touches the symbol of the block. Then the password contains at least one character from each set of four characters depicted on the symbol of each block.

E.g. User draws a line between the two blocks row 2, column 4 and row 4, column 1, the drawn line touches only the symbols of three blocks (row 2, column 4 and row 3, column 3 and row 4, column 1) Then the password comprises of at least one character from each of the three sets [2, ~, X,blank-space] and [3,;,blank-space,+] and [0,\,|,K] depicted on the blocks.

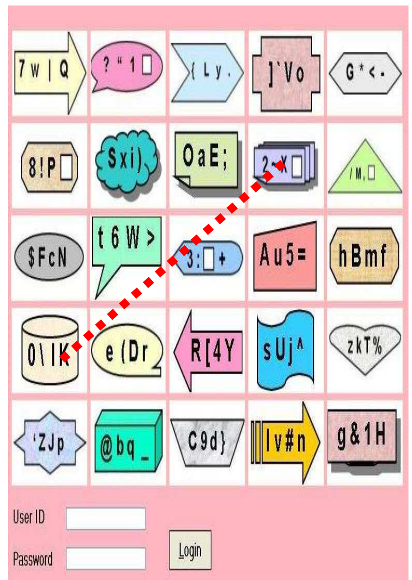


Fig 3 Line between non-adjacent blocks

Rule 3: If user draws a line on a single block, then the password contains at least one character on which the line passes.

E.g. If we draw a line on the characters 't ', '6 ',' w ' of block at row 3, column 2 then the password contains at least one character from the set [t,6,w].

Fig 4 Line on a single block

Rule 4: If the user draws the line on a single character of the symbol on a block then the password contains only that character.

E.g. If we draw a line on the character C of block at row 5, column 3 then the password is that character.

Fig 5 Line on single character of the symbol

IV. Usability Study & Security Analysis

We conducted a lab study with 23 participants out of which 15 were male and 8 were female. All the participants were post graduate students with their ages ranging from 22 to 26 years. A learning phase was conducted for practicing proposed graphical password scheme. They are given training initially explaining the concept of how to identify their password based on the rules proposed through the interface .The result was encouraging that novice users were able to identify the quadruplets formed with their password accurately. It took about 42 seconds on average to log in. Peeping attack is the attack where an attacker gets the secret information through direct observation when the user is entering his or her password. Alphanumeric systems are susceptible to peeping attack. In these attacks, typically the attacker gets a chance to observe the password entry for a short duration of time. As alphanumeric passwords are typically small, the attacker may see the secret by looking just for a while. On the other hand, peeping attack is not feasible against our proposed scheme as the user types or clicks on non password characters

V. Conclusion

Text-based authentication schemes are inherently insecure as they are subject to a tradeoff between usability and security, however they remain popular as their concept corresponds to an existing common model worldview making them an easy to understand concept. Graphical password has been designed to overcome the text-based password problems. Graphical passwords are more memorable compared to text-based passwords. In this paper, we proposed a new graphical password system resistant to peeping attack with promising usability features. The scheme provides a potential solution for the current problems faced by the other graphical password schemes. The proposed scheme provides larger password space than traditional text based passwords. This work can be extended by increasing the password space using more than three color character sets based upon user choice. The extension of the proposed schemes to hand-held mobile devices can be explored as future work.

REFERENCES

[1] 1. Abdullah et al (2008), Graphical Passwords: *Users' Affinity of Choice, AnAnalysis of Picture Attributes Selection*,IEEE 2008.

[2] R. N. Shepard. Recognition memory for words, sentences and pictures. *Journal of Verbal Learning and Verbal Behavior*, 6:156–163, 1967.

[3] R. U. Corporation. How the passface system works, 2005.

[4] D. Hong, S. Man, B. Hawes, and M. Mathews. A password scheme strongly resistant to spyware. In *in Proceedings ofInternational conference on security and management*, LasVergas, NV, 2002.

[5] I. Jermyn, A. Mayer, F. Monrose, M. K. Reiter, and A. D. Rubin. in proceedings of the design and analysis of graphicalpasswords. In *the 8th USENIX Security Symposium*, 1999.

[6] Renaud, K and De Angeli, A.My password is here! An investigation into visuo-spatial authentication mechanisms. 2004, Interacting with Computers, Vol. 16, pp. 1017-1041.

[7] Blonder, G E. Graphical Password. 5559961 United States, 1996.

[8] Wiedenbeck, S, et al. Authentication using graphical passwords:effect of tolerance and image choice. Pittsburgh, PA : ACM Press,2005. Proceedings of the 2005 symposium on Usable privacy and security (SOUPS). pp. 1-12.

[9] Jermyn, I, et al The Design and Analysis of Graphical Passwords.. Washington, D.C. USENIX Association, 1999. Proceedings of the 8th USENIX Security Symposium. pp. 1-14.

[10] Passfaces. [Online] Passfaces Coporation. [Cited: November 7, 2007.] http://www.passfaces.com.

[11] Dhamija, R., Hash visualization in user authentication The Hague,Netherlands : ACM Press, 2000. Conference on Human Factors in Computing System (CHI '00). pp. 279-280.

[12] Dhamija, R and Perrig, A. Déjà Vu: A user study using images for authentication. Denver, CO : USENIX Association, 2000. Proceedings of the 9th Conference on USENIX Security Symposium.

[13] Wiedenbeck, S, et al PassPoints: Design and longitudinal evaluation of a graphical password system.. 2005, 2005, International Journal of Human-Computer Studies, Vol. 63, pp. 102-127.

[14] Pierce, J D, et al Graphical Authentication: Justifications andObjectives Perth : Edith Cowan University, Western Australia, 2004. Proceedings of the 2nd Australian Information Security Management Conference. pp. 49-55.

[15] Pierce, J D, et al A conceptual model for graphical authentication Perth, Western Australia : Edith Cowan University, Western Australia, 2003. Proceedings of the 1st Australian Information Security Management Conference.

[16] De Angeli, A, et al Usability and user authentication: Pictorial passwords vs. PIN.. [ed.] P T McCabe. London : Taylor & Francis, 2003. Contemporary Ergonomics 2003. pp. 253-258.

[17] De Angeli, A, et al Is a picture really worth a thousand words? Exploring the feasibility of graphical authentication systems. 1-2, 2005, International Journal of Human-Computer Studies, Vol. 63, pp. 128-152.

[18] Jansen, W, et al. Picture password: a visual login technique for mobile devices. Department of Commerce, National Institute of Standard and Technology. Gaithersburg, MD : NISTIR, 2003.

[19] Jansen, W. [ed.] K Morgn and J M Spector ,"Authenticating mobile device users through image selection" 2004, In The Internet Society.[20] Authenticating Users on Handheld Devices. Jansen, W. 2003.Proceedings of the Canadian Information Technology Security Symposium.

[20] Hinds, C and Ekwueme, C. Winston-Salem, Increasing security and usability of computer systems with graphical passwords. NC : ACM Press, 2007. Proceedings of the 45th Annual Southeast Regional Conference. pp. 529-530.

[21] M. Shahid and M.A. Qadeer. Novel scheme for securing passwords. In Digital Ecosys-tems and Technologies, 2009. DEST'09. 3rd IEEE International Conference on, pages 223{227. IEEE, 2009.

[22] R. Jhawar, P. Inglesant, N. Courtois, and M.A. Sasse. Make mine a quadruple: Strengthening the security of graphical one-time pin authentication. In Network and System Security (NSS), 2011 5th International Conference on, pages 81{88. IEEE, 2011.

[23] Dictionary.com. Password | at dictionary.com, 2012. [Online; accessed 03-March-2012].

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 1, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

New Approach to Optimize the Time of Association Rules Extraction

Thabet Slimani
CS Department, Taif University, P.O.Box 888, 21974, KSA

Abstract
The knowledge discovery algorithms have become ineffective at the abundance of data and the need for fast algorithms or optimizing methods is required. To address this limitation, the objective of this work is to adapt a new method for optimizing the time of association rules extractions from large databases. Indeed, given a relational database (one relation) represented as a set of tuples, also called set of attributes, we transform the original database as a binary table (Bitmap table) containing binary numbers. Then, we use this Bitmap table to construct a data structure called Peano Tree stored as a binary file on which we apply a new algorithm called BF-ARM (extension of the well known Apriori algorithm). Since the database is loaded into a binary file, our proposed algorithm will traverse this file, and the processes of association rules extractions will be based on the file stored on disk. The BF-ARM algorithm is implemented and compared with Apriori, Apriori+ and RS-Rules+ algorithms. The evaluation process is based on three benchmarks (Mushroom, Car Evaluation and Adult). Our preliminary experimental results showed that our algorithm produces association rules with a minimum time compared to other algorithms.
Keywords: *Data Mining, Association Rules, Large Databases, Frequent Itemsets, Peano Trees (Ptree).*

1. Introduction

As a prominent tool for knowledge mining, Data mining [1] includes several techniques: Clustering, Association, Classification and Deviation. Knowledge Discovery in Data (KDD) constitutes an important advance in the area of data mining. It consists in the extraction of implicit knowledge (previously unknown and potentially useful), hidden in large databases. Association rule mining [2] is one of the principal problems treated in KDD and can be defined as extracting the interesting correlation and relation among huge amount of transactions. The task of association rule mining is to find interesting relationships from the data in the form of rules. The original application of association rule mining was on market basket analysis with the aim to study the buying habits of customers [3]. Currently, ARM has been the subject of several real-world applications in different areas requiring research groups of potential product or service, such as: medical diagnosis [4], biological database [5][6], electronic commerce [7][8] misuse detection [9].

Formally, an association rule is an implication relation in the form $X \rightarrow Y$ between two disjunctive sets of items X and Y. A typical example of an association rule on "market basket data" is that "80% of customers who purchase bread also purchase butter ". Each rule has two quality measurements, support and confidence. The rule $X \rightarrow Y$ has confidence c if c% of transactions in the set of transactions D that contains X also contains Y. The rule has a support S in the transaction set D if S% of transactions in D contain $X \cup Y$. The problem of mining association rules is to find all association rules that have a support and a confidence exceeding the user-specified threshold of minimum support (called MinSup) and threshold of minimum confidence (called MinConf) respectively.

Actually, frequent itemset mining and association rule mining became a wide research area in the field of data mining, and consequently a large number of quick and speed algorithms have been developed. The more efficient are those Apriori based algorithms or Apriori variations. The works that used Apriori as a basic search strategy, they also adapted the complete set of procedures and data structures [3][10][11]. Additionally, the scheme of this important algorithm was also used in sequential pattern mining [12], episode mining, functional dependency discovery & other data mining fields (hierarchical association rules [13]).

Another work was concentrated to develop faster algorithms for existing classical methods and adapting the algorithms into various states. As examples: multidimensional database mining [14], ontology based rule mining [15], association rule mining from the data cube [16][17][18], association rule mining in data warehouses [19], ontology based rule mining [20], parallel algorithms for association rule mining [21][22] and other algorithms.

Finding association rules is valuable for crossing-marketing [23] and attached mailing applications. Other applications include catalog design, add-on sales, store layout, and customer segmentation based on buying patterns. Besides application in the business area, mining association rule can also be applied to other areas, such as medical diagnosis [24], and remotely sensed imagery [25].

The databases involved in these applications are very large. Mining association rules in such databases may require substantial processing power. Therefore, it is necessary to have fast algorithms for this task. This observation motivates us to propose a new method for mining association rules in large databases. Given a relational database with various types of attributes (binary or not binary attributes), we first propose to convert the original database in a binary table (Bitmap). This transformation is a characteristic of the rough set method described in [26]. Next, we use a structure of data, called Peano tree (Ptree) which provides a lossless and compressed representation of Bitmap. By using Ptrees, an association rule mining algorithm with fast support calculation and significant pruning techniques are possible. The present work illustrates that using efficient data structures (Ptree) and our B-ARM algorithm (Binary Association Rule Mining) can be interesting for extracting the frequent itemsets[1] which is a time consuming task, especially when databases are large.

The Ptrees are used, in this context, to extend the Anding operation [27] of Ptrees to the attributes of the database. Using Ptrees, we do not make the expensive task of database scan each time we need to calculate the itemsets supports because, as we have already mentioned, concerned database is charged in a binary file, and Therefore we use the proposed B-ARM algorithm to extract the frequent itemsets on the specified Data Base in minimal time compared to other works.

This paper is organized as follows: in Section2, we describe briefly some related works. Section 3 summarizes the Ptree structure. Section 4 presents the specification of database attributes. In section 5, we details how to derive association rules using Ptrees. In Section 6, we describe our BF-ARM algorithm of the association rule mining. Details on implementation and experimental results are discussed in Section 7. Finally, we conclude with a summary of our approach and extensions of this work.

2. Literature Review

Early studies examined efficient mining association rules from different point of views. Apriori [28] is certainly the basic algorithm; it is developed for rule mining in large transaction databases. A DHP (Direct Hashing and Pruning) is an extension of the Apriori algorithm using a hashing technique [29]. A more recent algorithm called FDM (Fast Distributed Mining of association rules) was proposed by Cheung et al.[30], it is characterized by the generation of a small number of candidate sets and by the reduction of the number of messages to be passed at mining association rules. PincerSearch[31] spreads Apriori algorithm to generate the frequent itemsets. Depth-project [32] uses a dynamic reordering in order to reduce the research space. Another work realized by [33] proceeds to the improvement of the quality of the association rules by rough set technique. At least, nearer of our work, on one hand, FP-growth algorithm that represents the basis of transactions in the form of a compressed tree called FP-tree [34] and on the other hand, the MFItemsets algorithm (Maximum Frequent Itemsets) that represents the database as a truth table with an output Boolean function and sends back a body of Boolean products corresponding to the maximum frequent itemsets associated with the given transactions [35]. More recently, The work proposed by Rajalakshmi et al. [36] which identify maximal frequent itemsets based on minimum effort. The following paragraphs give a more detailed explanation of the previous approaches for more clarification:

Apriori: Apriori proposed by [28] is the fundamental algorithm. It searches for frequent itemset browsing the lattice of itemsets in breadth. The database is scanned at each level of lattice. Additionally, Apriori uses a pruning technique based on the properties of the itemsets, which are: If an itemset is frequent, all its sub-sets are frequent and not need to be considered.

DHP: DHP algorithm (Direct Haching and Pruning) proposed by [29] is an extension of the Apriori algorithm, which use the hashing technique with the attempts to efficiently generate large itemsets and reduces the transaction database size. Any transaction that does not contain any frequent k-itemsets cannot contain any frequent (k+1)-itemsets and such a transaction may be marked or removed.

FDM: FDM (Fast Distributed Mining of association rules) has been proposed by [30], which has the following distinct features.

1. The generation of candidate sets is in the same spirit of Apriori. However, some relationships between locally large sets and globally large ones are explored to generate a smaller set of candidate sets at each iteration and thus reduce the number of messages to be passed.
2. The second step uses two pruning techniques, local pruning and global pruning to prune away some candidate sets at each individual sites.
3. In order to determine whether a candidate set is large, this algorithm requires only O(n) messages for support count exchange, where n is the number of sites in the network. This is much less

[1] Itemsets which have support above the user-specified minimum support.

than a straight adaptation of Apriori, which requires $O(n^2)$ messages.

PincerSearch: The Pincer-search algorithm [31] proposes a new approach for mining maximal frequent itemset which combines both bottom-up and top-down searches to identify frequent itemsets effectively. It classifies the data source into three classes as frequent, infrequent, and unclassified data. Bottom-up approach is the same as Apriori. Top-down search uses a new set called Maximum-Frequent-Candidate-Set (MFCS). It also uses another set called the Maximum Frequent Set (MFS) which contains all the maximal frequent itemsets identified during the process. Any itemset that is classified as infrequent in bottom-up approach is used to update MFCS. Any itemset that is classified as frequent in the top-down approach is used to reduce the number of candidates in the bottom–up approach. When the process terminates, both MFCS and MFS are equal. This algorithm involves more data source scans in the case of sparse data sources.

Depth-project: DepthProject proposed by Agarwal et al., (2000) [32] also mines only maximal frequent itemsets. It performs a mixed depth-first and breadth-first traversal of the itemset lattice. In the algorithm, both subset infrequency pruning and superset frequency pruning are used. The database is represented as a bitmap. Each row in the bitmap is a bitvector corresponding to a transaction and each column corresponds to an item. The number of rows is equal to the number of transactions, and the number of columns is equal to the number of items. By using the carefully designed counting methods, the algorithm significantly reduces the cost for finding the support counts.

FP-tree : FP-tree proposed by Han et al., (2000) [34] is a compact data structure that represents the data set in tree form. Each transaction is read and then mapped onto a path in the FP-tree. This is done until all transactions have been read. Different transactions that have common subsets allow the tree to remain compact because their paths overlap. the size of the FP-tree will be only a single branch of nodes. The worst case scenario occurs when every transaction has a unique itemset and so the space needed to store the tree is greater than the space used to store the original data set because the FP-tree requires additional space to store pointers between nodes and also the counters for each item.

GenMax: GenMax proposed by Gouda and Zaki, [37] a backtrack search based algorithm for mining maximal frequent itemsets. GenMax uses a number of optimizations to prune the search space. It uses a novel technique called progressive focusing to perform maximality checking, and diffset propagation to perform fast frequency computation.

FPMax: FPMax (Frequent Maximal Item Set) is an algorithm proposed by Grahne and Zhu, (2005) [38] based on FP Tree. It receives a set of transactional data items from relational data model, two interesting measures Min Support, Min Confidence and then generates Frequent Item Sets with the help of FPTree. During the process of generating Frequent Item Sets, it uses array based structure than tree structure. Additionally, the FPMax is a variation of the FP-growth method, for mining maximal frequent item sets. Since FPMax is a depth-first algorithm, a frequent item set can be a subset only of an already discovered MFI.

Method based on minimum effort: The work proposed by Rajalakshmi et al. (2011) [36] describes a novel method to generate the maximal frequent itemsets with minimum effort. Instead of generating candidates for determining maximal frequent itemsets as done in other methods [31], this method uses the concept of partitioning the data source into segments and then mining the segments for maximal frequent itemsets. Additionally, it reduces the number of scans over the transactional data source to only two. Moreover, the time spent for candidate generation is eliminated. This algorithm involves the following steps to determine the MFS from a data source:

1. Segmentation of the transactional data source.
2. Prioritization of the segments
3. Mining of segments

3. Ptree Structure

The data structure tree Peano (Ptree), also called "peano count tree" is a compact and efficient representation used to store a database (originally an image) as a binary bits (0 and 1). This structure was initially introduced for the representation of spatial data such as RSI data applications (Remotely Sensed imagery) [39][27]. Using Ptree structure, all the count information can be calculated quickly. This facilitates efficient ways for data mining.

A Ptree is a quadrant based tree. The Ptree principle is to divide, recursively, the totality of spatial data into quadrants and counting the bits having the value "1" for each quadrant, thus forming a computation quadrants tree. In figure 1, 55 is the number of bits in one complete picture, the root level is labeled level 0. The numbers at the next levels (level 1) are, 16, 8, 15 and 16, are the 1-bit counts for the four major quadrants. The quadrants composed entirely of 1-bits are called a "pure 1 quadrant" (the first and last quadrant with 16 value are a pure 1 quadrant and we do not need sub-trees for these two quadrants, so these branches terminate) and similarly, the quadrants composed entirely of 0-bits are called a "pure 0 quadrant" (which also terminate). This

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 1, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

process is repeated recursively using the Z-ordering of the four sub-quadrants at each new level. Eventually, every branch terminates in the leaf level or when each quadrant is a pure quadrant.

The Ptrees are similar in their construction to other existing data structures, for example Quadtrees (Samet 1984) and HHcodes[1]. The similarities between Ptrees, quadtrees, and HHcodes[2] are that they are quadrant based. The difference is that Ptree include occurrence counts. Trees are not indexed, but they are representations of the dataset itself.

When using the Ptree structure, any information calculation can be completed very fast. The performance analysis realized in [27] shows that Ptree produces a good cost computation (CPU time) and reduce the storage space compared to the original data.

Peano mask tree (pm-tree) is a variation of the ptree data structure. Pm-tree is a similar structure in which masks rather than counts are used. In pm-tree structure, to represent pure-1, pure-0 and mixed quadrant we use a 3-value logic. In pm-tree structure, to represent pure-1, pure-0 and mixed quadrant we use a 3-value logic. Pm-tree is helpful for the optimization of anding operation between two ptrees.

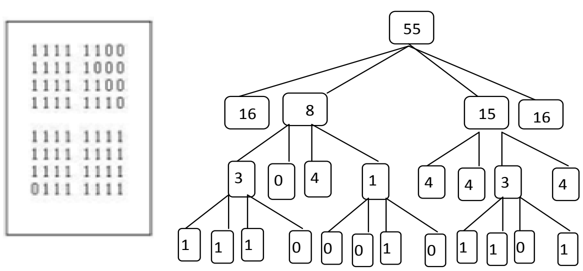


Figure1. Ptree for 8*8 image

A PM-tree example is given in the Figure 2. Other variations can be used, such as P1-tree and P0-Tree. In P1-tree, we use 1 to indicate the pure-1 quadrant while use 0 to indicate others. In P0-tree, we use 1 to indicate the pure-0 quadrant while use 0 to indicate others. Both P1-tree and P0-tree are lossless representations of the original data [39].

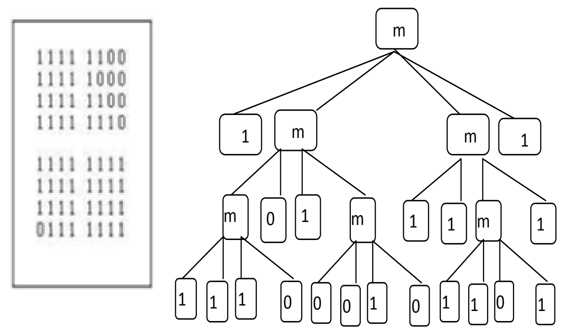


Figure1. PMtree for 8*8 image

4. Specification of Database Attributes

A Database is represented by a binary or bitmap table whose columns are attributes and each attribute owns a limited set of values (items) known by the domain attributes of the database. A database can have two types of attributes domain: Binary attributes domain (BAD) and non-binary attributes domain (NBAD).

Binary Attributes domain: A Binary attributes domain is represented by a vector $\vec{v} \subset \{v1, v2\}$ with size k, such that the values v1 and v2 are taken from the set {0,1}, and k is the number of k-tuples of values taken from {0, 1}. 1-tuple represents a tuple of the database or transaction in terms of the market basket data analysis.

A database A^n n-dimensional is constituted of n binary vectors, when each vector has 2^n size and is constituted in turn with 4 binary vectors $2^n/4$ (for simplicity reasons, we decompose each binary vector into four quadrants).

A^n Lines (transactions) repesent all combinations of n possible binary values 0 and 1. In the example given in table.1, the presence of a computer item in a transaction or its absence represents its domain {purchased, not purchased} and the binary transformation makes the attribute value a1 =1, if the computer is purchased or a1 = 0 if the computer is not purchased.

Table1. The transformation of raw data into a bitmap representation for BAD.

Tid	Computer		Tid	a1
1	Purchased		1	1
2	Not purchased	➔	2	0
3	Purchased		3	1
4	Not purchased		4	0
….	….		…..	…..

[1] http ://www.statkart.no/nlhdb/iveher/hhtext.html
[2] http ://www.statkart.no/nlhdb/iveher/hhtext.html

Non Binary Attributes domain: A non Binary attribute domain Aj is consituted with j items of the Database and represented by $\sum_{i=1}^{n} j * i$ binary vectors where n is the number of attributes of the non binary attributes domain. For example, for a better representation of the benefit of a client, we associated to the attribute "income" the domain with three (j=3) items {high, medium, low} defined as follows: a1 = "high income" a2 = "middle income" and a3 = "low income" and represented by the following binary table (Table 2):

Table2. The transformation of raw data into a bitmap representation for NBAD

Tid	Income		Tid	a1	a2	a3
1	High		1	1	0	0
2	Meduim	➔	2	0	1	0
3	Low		3	0	0	1
4	High		4	1	0	0
….	……….		…..	…	…	…

5. Association Rule Mining using Ptree

Given a user-specified minimum support and minimum confidence, the problem of mining association rules is to find all the association rules whose support and confidence are larger than the respective thresholds specified. Thus, it can be decomposed into two subproblems :

Finding the frequent itemsets which have support above the user-specified minimum support.

Deriving all rules, based on each frequent itemset, which have more than user-specified minimum confidence.

The whole performance is mainly determined by the first step, which is the generation of frequent itemsets. Once the frequent itemsets have been generated, it's straightforward to derive the rules. To solve this problem and to improve the performance, Apriori algorithm was proposed [28]. Apriori algorithm generates the candidate itemsets to be counted in the pass by using only the itemsets found large in the previous pass - without considering the transactions in the database.

The key idea of Apriori algorithm lies in the "downward-closed" property of support which means if an itemset has minimum support, then all its subsets also have minimum support. An itemset having minimum support is called frequent itemset (also called large itemset). So any subset of a frequent itemset must also be frequent. The candidate itemsets having k items can be generated by joining frequent itemsets having k-1 items, and deleting those that contain any subset that is not frequent.

Start by finding all frequent 1-itemsets (itemsets with 1 item); then consider 2-itemsets, and so forth. During each iteration only candidates found to be frequent in the previous iteration are used to generate a new candidate set during the next iteration. The algorithm terminates when there are no frequent k-itemsets. Since finding frequent itemsets is a time consuming task, especially when databases are large, we use a Bitmap table (containing binary data) to organize original database and the Peano tree (Ptree) structure to represent Bitmap tables in a spatial data mining-ready-way. Ptrees are a lossless representation of the original database.

5.1. Model Representation

The rough set method [26] operates on data matrices, called "information tables" which contain data about the universe Ω of interest, condition attributes Ωc and decision attributes Ω d. The goal is to derive rules that give information how the decision attributes depend on the condition attributes.

Let us consider the original database represented by the corresponding information contained in Table 3 where condition attributes are {A, B, C, D, E, F} associated to the following products {cartridge printer, video reader, car, computer, movie camera, printer} and decision attribute is {G} corresponding to {graphic software}. Each attribute has two non null values {yes, no}. So, there are seven items (bitmap-attributes) for the resulting bitmap-table {A, B, C, D, E, F and G}.

Table 3. Original table (database) with its equivalent Bitmap table.

T_{id}	A	B	C	D	E	F	G
T_{id1}	yes	yes	no	no	no	no	no
T_{id2}	yes	yes	yes	yes	yes	yes	no
T_{id3}	no	yes	no	yes	no	no	yes
T_{id4}	no	yes	no	no	yes	no	yes
T_{id5}	no	no	no	yes	no	yes	yes
T_{id6}	no	no	no	yes	yes	no	yes
T_{id7}	no	yes	no	no	yes	no	no
T_{id8}	no	yes	no	yes	yes	yes	no

Tid	A	B	C	D	E	F	G
T_{id1}	1	1	0	0	0	0	0
T_{id2}	1	1	1	1	1	1	0
T_{id3}	0	1	0	1	0	0	1
T_{id4}	0	1	0	0	1	0	1
T_{id5}	0	0	0	1	0	1	1
T_{id6}	0	0	0	1	1	0	1
T_{id7}	0	1	0	0	1	0	0
T_{id8}	0	1	0	1	1	1	0

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 1, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

In the first step, all possible rules are constructed from all bitmap-attributes of the table. All rules not fulfilling the minimum support (MinSup=10%) and minimum confidence (MinConf=30%) should be deleted.

5.2. From Attributes to Ptree

According Ptree representation, the set of tuples must be a power of 2. The number of the tuples in a database is transformed to the nearest power of 2, knowing than the completed tuples are itemsets uniquely including the value 0. In Ptree approach, each column of Bitmap table is represented by a vector of bits of which cuts it divisible by 4, called basic Ptree.

For simplicity reasons, we suppose that the fan-out is four. For every vector of bits, a basic Ptree is associated. There are six basic Ptrees for the universe Ω_c of Table.2, and since each Ptree presents a number of bits divisible by 4, therefore, it is constituted by four under quadrants of which the origin quadrant is the entirety of the bits forming Bitmap table. As the header of Ptree files contains the root count, the root counts of Ptrees are immediately accessible, and will conveniently replace the necessity of using original data to count the number of transactions containing candidate frequent itemsets.

Ptree_A Ptree_B …………… Ptree_G

Tid	A	B	C	D	E	F	G
T_{id1}	1	1	0	0	0	0	0
T_{id2}	1	1	1	1	1	1	0
T_{id3}	0	1	0	1	0	0	1
T_{id4}	0	1	0	0	1	0	1
T_{id5}	0	0	0	1	0	1	1
T_{id6}	0	0	0	1	1	0	1
T_{id7}	0	1	0	0	1	0	0
T_{id8}	0	1	0	1	1	1	0

Figure 1. Tabular Representation of Ptrees.

In a similar way to the generation of the sequence of Peano from the spatial data, we create the Ptree in a Bottom-up way. The generation of Ptree depends on the number of fan-out in the internal nodes of the Ptree and in the root node. To represent Ptree with different fans-outs, we introduce the Ptree-(r-i) notation ; where r = the fan-out of the root node and i = the fan-out of all the internal nodes of the level 1. We adopt in our work, the representation Ptree-(4-4-n), it means that this structure divides the database tuples into 4 blocks (the block of a transaction must have at minimum 4 tuples). For example, if the number of tuples is less than 16, one completes by 0 to obtain the Ptree format on 16 tuples. Generally, if the number of transactions is inferior to 2^{n+1} and superior to 2^n then the basic Ptree is stored with 2^{n+1} number.

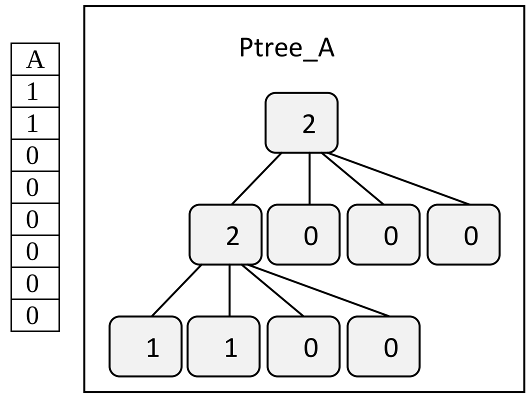


Figure 2. Snaphsot of Ptrees representation (Ptree_A).

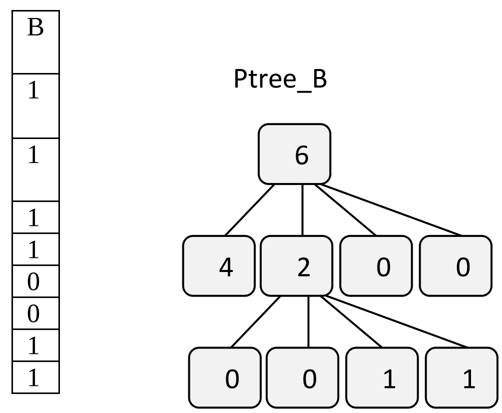


Figure 3. Snaphsot of Ptrees representation (Ptree_B).

5.3. Ptree Anding Operation

ANDing is a very important and frequently used operation for Ptrees. There are several ways to execute Ptree ANDing operation. We can execute anding level-by-level starting from the root level. Table. 4 gives the rules for performing Ptree ANDing. Operand 1 and Operand 2 are two ptrees with root X_1 and X_2 respectively. Using PM-trees, X_1 and X_2 could be any value between 1, 0 and m (3-value logic representing pure-1, pure-0 and mixed quadrant). For example, a pure-1 Ptree combined with any ptree will have as consequence the second operand and a pure-0 ptree with any ptree will result in the pure-0 ptree**.**

Table 4. Ptree anding rules.

Operand 1	Operand 2	Result
1	X_2	Subtree with root X_2
0	X_2	0
X_1	1	Subtree with root X_1
X_1	0	0
m	m	0 if four sub-quadrants result in 0 value; m otherwise

6. Algorithm BF-ARM (Binary File Association Rule Mining)

In accordance with the Apriori algorithm, we wish to find all frequent 1-itemsets first. We do not need to scan the entire data set and employ a counter for each item. We only need to access the root count of the Ptree for every item in the table. Then a simple calculation will indicate whether the item is a frequent 1-itemset or not. If the root count of an item's Ptree divided by the total number of transactions is greater than Minsup, then the item is a frequent 1-itemset.

Using Ptrees structure has now saved one scan of the entire data set, along with the necessity of memory buffer management. Next we want to find frequent 2-itemsets. The only possible candidate 2-itemsets are made up of frequent 1-itemsets. All other Ptrees will be ignored in finding frequent 2-itemsets, so we are working with a subset of the original data set. The candidate 2-itemsets are all the pairwise combinations of frequent 1-itemsets. For each candidate 2-itemset, the Ptrees for the two items are ANDed to produce a third, derived Ptree that represents the presence of both items in transactions. Note that the ANDing process is fast enough that there is no need to save the new Ptree for further operations (Ding et al, 2002). To determine support for any candidate 2-itemset, we divide the root count of the new, 2-itemset Ptree by the number of transactions in the table. If support is greater than Minsup, we'll calculate the confidence levels to test against Minsup. The numbers needed to calculate confidence are, once again, the root counts of Ptrees.

Now, knowing the frequent 2-itemsets we can continue with discovering frequent 3-itemsets from candidate 3-itemsets and so on until no candidate k-itemsets exist. The basic algorithm is to AND all the Ptrees of the items in the candidate k-itemset, divide the root count of the new Ptree by the total number of transactions, and test for Minsup. In every case, there is no need to scan the data since the necessary counts exist already within the Ptrees derived from AND operations.

Algorithm ***BF-ARM***
Data Discretization
Ptrees_Storage
For each attribute $i \subset \Omega_c$
  $C_1 = F_1$
End For
$C_k = C_1$
Do While ($C_k \neq \emptyset$)
  For each attribute $i \subset C_k$
    For each attribute $j \subset \Omega_d$
    F_{ij} = AND_Ptreebase(i,j)
    Storage_Ptrees
    End For
   $F_k = F_k \cup F_i$ //itemsets candidats
  End For
 C_k = Fk{c ∈ Ck | c.count ≥ MinSup}
End While

Function Storage_Ptrees
 For (bandj=1; j<I; j++)
  root[j] :=rootcount(1 ; bandj)
 //vector storing the roots of Ptrees.
  If($2^n \le N_t$ and $N_t < 2_{n+1)}$ **then**
   For (i :=N_t ; i≤ 2^{n+1} ; i++)
    bandj[i+1] :=0 ;
   End For
  End If
  k :=0 ;
 For (i :=1 ; i≤2^n; 2^n/4)
  k :=k+1 ;
  rootsBandj[k] :=rootcount (1 ; bandj) ;
  End For
  For (i :=1 ; i≤2^n; 2^n/4)
   If (rootsBandj[i]<>2^n or rootsBandj[i]<>0)
   then
    bitsBandj[i] :=rootsBandj[i] ;
  // bits vectors
   End If
  End For
End For

The *rootcount* function, used in the storage procedure of Ptree, is to calculate admissible itemsets counts directly by ANDing the appropriate basic Ptrees instead of scanning the original databases. Let N_t be the number of tuples; n is initialized to 3 and I be the total number of attributes.

The list of the strong rules generated in the example described previously is summarized in Table 5. The rules are classified by decision attribute G or F.

Table5. Strong rules generated by the BF-ARM algorithm.

	Decision Attribute = G	Decision Attribute = F
Strong rules with 2 attributes	B→G D→G E→G F→G	A→F B→F D→F E→F
Strong rules with 3 attributes	B,D→G	A, B→F A,C→F A,D→F A,E→F B,C→F B,D→F C,D→F
Strong rules with 4 attributes		A,B,C→F A,B,D→F A,B,E→F A,C,D→F B,C,D→F B,C,E→F B,D,E→F C,D,E→F
Strong rules with 5 attributes		A,B,C,D→F A,B,C,E→F A,C,D,E→F B,C,D,E→F
Strong rules with 6 attributes		A, B,C,D,E→F

7. Experimental Results

In this section, we compare our work with the two approaches Apr+ (Hybrid Association Rule Algorithm Apriori+) and Rs+ (Rough Set Based Rule Generation Algorithm RS-rules+) (Delic et al. 2002). The procedure Apr+ is an extension of the method "faster association rule" combined with the procedure "rough set". The derived rules are produced on the basis of the successive reduction of the useless rules. If there is a given fixed decision attribute, all itemsets without this attribute can be ignored for rule generation. Besides the Apr+ procedure of rule generation with a fixed decision attribute, the procedure Rs+ offers the possibility of varying the selected decision attributes, so, each attribute can be included either as a decision or condition attribute. In our work, we use the principle for deriving rules with a fixed decision attribute, but we add the notion of the Ptree structure to accelerate the processor time generation of the strong rules. Furthermore, in our work, we don't produce redundant rules, while in Rs+ and Apr+, redundant rules are produced and removed by the continuation.

The comparison of our work is facilitated by the use of a benchmark data set[1] concerning Car Evaluation Database, Mushroom Database and Adult database. For example, the car evaluation database contains 1728 tuples and 25 values of attributes (items) in the Bitmap table.

The preliminary experimental results given in Table 6 show that the computing times were in favor of our BF-ARM algorithm (Figure4). Indeed, bases on Car Evaluation benchmark, our work produces a CPU time to generate strong rules equal to 0,083 min for fixed decision attributes and a CPU time equal to 0,067 min for no fixed decision attributes. By running Apr+, Rs+, and BF-ARM algorithms, we got identical rules.

Table 6. Comparative table for the algorithms RS-ules+ (Rs+), Apriori+ (Apr+), Apriori (APR) and BF-ARM.

Database	**Car Evaluation CPU Time(Min)**		**Mushroom CPU Time(Min)**		**Adult CPU Time(Min)**	
MinSupp	10%		35%		17%	
MinConf	75%		90%		94%	
	Fixed Decision Attribute					
Method	Yes	No	Yes	No	Yes	No
RS+	1.15	3.15	3.32	15	64	233
APR+	1.12	1.12	2.02	2.02	44	44
APR		1.10		2		44
BF-ARM	0.083	0.067	0.3	0.223	6.29	4.54

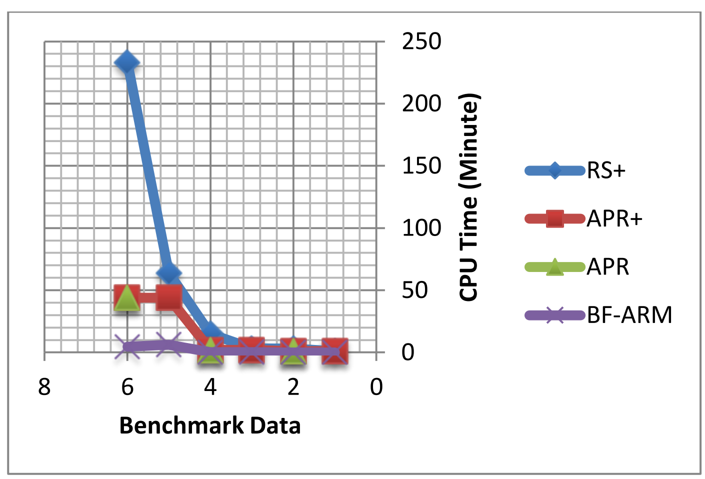


Figure 4. Comparative Snapshot for the algorithms RS-rules+ (Rs+), Apriori+ (Apr+), Apriori (APR) and B-ARM.

8. Conclusion

In this paper, we propose a new method to derive association rules from large databases using Ptree structure. The Ptree structure is a space efficient, lossless, data mining ready structure for binary datasets. The adopted association rules method specified by our algorithm BF-ARM is based on the concept of pruning by minimum support and minimum confidence directly to produce strong association rules. The discovery of similarities between attributes/items was based on the rules of Ptrees ANDing. Based on the benchmark data, we compared the quality of the produced rules and the necessary computing times of the algorithms. It turned out the generated rules of Apr+, Rs+, Apr and BF-ARM does not differ. But, the computing times were in favor of BF-ARM algorithm. Our work is beneficial because on the one hand, it avoids the direct scanning of database (an expensive operation in memory and computation time), which greatly exceeds the capacity of computers, despite their rapid evolutions and, secondly, it provides a gain of attributes comparison because the comparison is performed by a block of tuples. Another interesting direction is the extension of our association rule mining method by adding time constraints (new area for sequence identification).

References

[1] K.J.Cios, W. Pedrycz, R.W. Swiniarski, & L.A.Kurgan. *Data mining: A knowledge discovery approach*. New York, NY: 2012, Springer.

[2] X. Liu, K. Zhai, &, W. Pedrycz. An improved association rules mining method. *Expert Systems with Applications*. 2012, *39*(1), pp.1362–1374. doi:10.1016/j. eswa.2011.08.018.

[3] R. Agrawal, R. Imielinski, & A. Swami. Mining associations between sets of items in massive databases. In *Proceedings of the ACM SIGMOD Conference on Management of Data*, Washington, 1993, pp. 207-216.

[4] Z.Xing, & J. Pei. Exploring disease association from the NHANES data: Data mining, pattern summarization and visual analysis. *International Journal of Data Warehousing and Mining*, 2010, *6*(3), 11–27. doi:10.4018/jdwm.2010070102.

[5] A. Rodriguez, J.M.Carazo, & O. Trelles. Mining association rules from biological databases. Journal of the American Society for Information Science and Technology, 2005, 56 (5), **493-504.**

[6] R.Martinez, N. Pasquier, & C. Pasquier. GenMiner: Mining non-redundant association rules from integrated gene expression data and annotations. *Bioinformatics (Oxford, England)*, 2008, *24*(22), 2643–2644.doi:10.1093/bioinformatics/btn490 PMID:18799482.

[7] R. Natarajan, & B. Sheka. Interestingness of association rules in data mining: Issues relevant to e-commerce. 2005, *Sadhana*, *30*, 291–309. doi:10.1007/ BF02706249.

[8] Z. Yiyang, J. Jianxin. "An associative classification-based recommendation system for personalization in B2C ecommerce applications", Expert Systems with Applications, 2007, 33 ,357–367.
[9] K.Keerthi, P.Sreenivas. Reducing Network Intrusion Detection using Association rule and Classification algorithms. *International Journal of Computer Trends and Technology,2012, 3(6), 752-756.*

[10] S. Brin, R. Motwani, J.D. Vllman & S. Tsur. "Dynamic itemset counting and implication rules for market basket data", SIGMOD Record (ACM Special Interest Group on Management of Data),1997, 26(2), 255.

[11] H.Toivonen. "Sampling large databases for association rules", In the VLDB '96 Proceedings of the 22th International Conference on Very Large Data Bases, pp.134-145, 1996.

[12] R. Agrawal and R. Srikant. "Mining sequential patterns" In P.S.Yu and A.L.P. Chen, editors, Proc.11the Int. Conf. Data engineering. ICDE, pp. 3-14. IEEE, 1995, pp, 6-10.

[13] Y.F.Jiawei Han. "Discovery of multiple-level association rules from large databases" In Proc. of the 21St International Conference on Very Large Databases (VLDB), Zurich, Switzerland. 1995, Pp. 420-431.

[14] J. Liu, H. Wang, & H. Zhou. The application of college employing management system based on improved multi-dimension association rule mining algorithm. In *Proceedings of the International Conference on Intelligent Human-Machine Systems and Cybernetics*, China (pp. 135-137), 2011.

[15] P. Manda, S. Ozkan, H. Wang, F. McCarthy, S.M . Bridges. Cross-Ontology Multi-level Association Rule Mining in the Gene Ontology. PLoS ONE 7(10): e47411, 2012. doi:10.1371/journal.pone.0047411.

[16] R. Messaoud, R. S. Loudcher, O. Boussaid, & R.Missaoui. Enhanced mining of association rules from data cubes. In *Proceedings of the 9th ACM International Workshop on Data Warehousing and OLAP*, Arlington, 2006, pp. 11-18.

[17] R.B. Messaoud, S.L.Rabaseda, R.Missaoui, and O.Boussaid, OLEMAR: An Online Environment for Mining Association Rules in Multidimensional Data. pp. 1–35 *in*: D. Taniar (ed.), Data Mining and Knowledge Discovery Technologies. IGI Publishing, Hershey, New York, 2008.

[18] J.J.Jigna, & P.Mahesh. Association rule mining method on OLAP cube. *International Journal of Engineering Research and Applications*, *2*(2), 1147–1151, 2012.

[19] H. C. Tjioe, & D.Taniar. Mining Association Rules in Data Warehouses.*International Journal of Data Warehousing and Mining (IJDWM)*, *1*(3), 28-62, 2005. doi:10.4018/jdwm.2005070103.

[20] S. X. Xia, F.Li, L.Zhang."Ontology-Based Association Rule Quality Evaluation Using Information Theory," in International Conference on Computational and Information Sciences, Chengdu, China, 2010, pp.170-173.

[21] Y. Junrui, Y. Yashuang. "A parallel algorithm for mining association rules," *2nd International Conference on Networking and Digital Society (ICNDS), 2010*, 30-31 May 2010. Wenzhou, China (pp.475-478) doi: 10.1109/ICNDS.2010.5479242.

[22] G. G. Zhang, C. Z. Xu, P. C.-Y. Sheu, & H. Yamaguchi. Parallel association rule mining for medical applications. In *Proceedings of the IEEE 11th International Conference on Bioinformatics and Bioengineering*, Beijing, China, 2011,pp. 148-154.

[23] K.Wang., S. Zhou, & G. Webb. Mining customer value: from association rules to direct marketing, In Proceedings of the IEEE International Conference on Data Engineering, 2005, vol 11, pp.57-80.

[24] S. Doddi, A.Marathe, S.S. Ravi & D.C. Torney. Discovery of association rules in medical data. Journal of Informatics for Health and Social Care 2001 26:1, 25-33.

[25] J. Dong, W. Perrizo, Q. Ding, & J. Zhou. The application of association rule mining to remotely sensed data. In *Proceedings of the 2000 ACM symposium on Applied computing - Volume 1* (SAC '00), Janice Carroll, Ernesto Damiani, Hisham Haddad, and Dave Oppenheim (Eds.), 2000, Vol. 1. ACM, 340-345. DOI=10.1145/335603.335786.

[26] T.Munkata. Rough Sets. In Fundamentals of the New Artificial Intelligence, New York : Springer-Verlag, 1998, pp. 140-182.

[27] Q. Ding, M. Khan, A. Roy & W. Perrizo. The Ptree Algebra. In Proc. of ACM Symposium on Applied Computing (SAC'02), 2002, Madrid, Spain, pp. 413-417.

[28] R. Agrawal & R. Srikant. Fast Algorithms for Mining Association Rules. In Proc. 20th Int. Conf. Very Large Data Bases (VLDB), 1994, pp. 487-499.

[29] J.S. Park, M.S. Chen & P.S. Yu. An Effective Hash-based Algorithm for Mining Association Rules. In Proc. 1995 ACM SIGMOD International Conference on Management of Data, 1995, pp. 175-186.

[30] C.Cheung, J. Han, V.T. Ng, A.W. Fu & Y. Fu. A Fast Distributed Algorithm for Mining Association Rules. In Proc. of 1996 Int'l Conf. on Parallel and Distributed Information Systems (PDIS'96), 1996, Miami Beach, Florida, USA.

[31] D.Lin & Z. M.Kedem. Pincer Search : A New Algorithm for Discovering the Maximum Frequent Set. In Proc. Int. Conf. on Extending Database Technology, 1998.

[32] R.C. Agarwal, C.C. Aggarwal & V.V.V. Prasad Depth First Generation of Long Patterns. In Proc. of the 6th Int. Conf. on Knowledge Discovery and Data Mining, 2000, pp. 108-118.

[33] D.Delic, L.Lenz & N.Neiling .Improving the Quality of Association Rule Mining by means of Rough Sets. Free university of Berlin, Institute of Applied Computer Science, Garystr. 21, D-14195, 2002, Berlin, Germany.

[34] J. Han, J. Pei & Y. Yin. Mining Frequent Patterns without Candidate Generation. In Proc. 2000 ACM SIGMOD Intl. Conference on Management of Data.

[35] A. Salleb & Z. Maazouzi. Approche Booléenne pour l'extraction des itemsets fréquents maximaux. In Conf. d'Apprentissage (CAp'02), 2002, Orléans, pp. 111-122.

[36] M. Rajalakshmi, T. Purusothaman, R Nedunchezhian. International Journal of Database Management Systems (IJDMS), Vol.3, No.3, August 2011, pp. 19-32, 2011.

[37] K.Gouda, and M.J.Zaki, 'GenMax : An Efficient Algorithm for Mining Maximal Frequent Itemsets', Data Mining and Knowledge Discovery, 2005, Vol 11, pp. 1-20.

[38] G.Grahne and G.Zhu. 'Fast Algorithms for frequent itemset mining using FP-trees', in IEEE transactions on knowledge and Data engineering, 2005, Vol 17, No. 10, pp. 1347-1362.

[39] W. Perrizo, Q. Ding, Q. Ding & A. Roy. On Mining Satellite and other Remotely Sensed Images. In Proc. of Workshop on Research Issues on Data Mining and Knowledge Discovery, 2001, pp. 33-44.

Adaptive Neuro-Fuzzy Inference System (ANFIS) Based Software Evaluation

Khyati M. Mewada [1], Amit Sinhal[2] and Bhupendra Verma[3]

**[1] Dept. of Information Technology, Technocrats Institute of Technology
Bhopal, Madhya Pradesh, India**

**[2] Dept. of Computer Sc. & Engg., Technocrats Institute of Technology
Bhopal, Madhya Pradesh, India**

**[3] Dept. of Dept. of Computer Sc. & Engg., Technocrats Institute of Technology (Excellence)
Bhopal, Madhya Pradesh, India**

Abstract

Software metric is a measure of some property of a piece of software or its specifications. The goal is to obtain reproducible and quantifiable measurements, which may have several valuable applications in schedule and budget planning, effort and cost evaluation, quality assurance testing, software debugging, software performance optimization, and optimal personnel task assignments. Software effort evaluation is one of the most essential and crucial part of software project planning for which efficient effort metrics is required. Software effort evaluation is followed by software cost evaluation which is helpful for both customers and developers. Thus, efficiency of effort component of software is very essential. The algorithmic models are weak in estimating early effort evaluation with regards to uncertainty and imprecision in software projects. To overcome this problem, there are various machine learning methods. One of the methods is soft computing in which there are various methodologies viz., Artificial Neural Network, Fuzzy Logic, Evolutionary computation based Genetic Algorithm and Meta-heuristic based Particle Swarm Optimization. These methods are good at solving real-world ambiguities. This paper highlights the design of an efficient software effort evaluation model using Adaptive Neuro-Fuzzy Inference System (ANFIS) for uncertain datasets and it shows that this technique significantly outperforms with sufficient results.

Keywords: *Software metrics, Software effort evaluation, Cost evaluation, Soft Computing Techniques, COCOMO, ANFIS.*

1. Introduction

The goal of any successful software project is to develop quality software within time, cost and resource constraints. This can be achieved consistently, only through effective management of the software development process. Well-defined measures of the process and the product are necessary to exercise control and to bring about improvement in the software development process. Software metrics are quantitative measures that provide the basis for effective management of the software development process. Software metrics are used to improve software productivity and quality. “Metric is a quantitative measure of the degree to which a given attribute is possessed by a system or its component or by a process.” Software metrics are measures that are used to quantify different attributes of a software product, software development resource and software development process. *Software metrics* deals with the evaluation and measurement of different attributes of the software product and the software development process [1]. There are three kinds of software metrics: procedure metrics, project metrics, and product metrics [2].

- ✓ Procedure metrics measure the resources (time and cost) that a program development effort will take. They are useful for the administration and management of the project.
- ✓ Project metrics give information about the actual situation of the project. These metrics include costs, effort, risks, and quality. These are used to improve the development process of the project.
- ✓ Product metrics assess quality information about the program. These metrics focus on reliability, maintainability, complexity, and reusability of all or part of the software developed for the program.

The reliability of these software metrics as predictors bugs has been studied and tested by many researchers [3, 4, 5], who have used different regression models applied to different languages. All of these researchers have claimed software metrics to have good capabilities as indicators of bugs.

Metrics are seen as force multipliers in improvement initiatives and quality movements. Metrics have led organizations and individuals in a process of self-discovery of goals, capabilities and constraints. Inspired by metrics, data patterns, evaluation models for bug fixing have been constructed and as a result the bug evaluation task has been refined and redefined in many organizations. The most vital contribution of metrics is the decision-making support. Constant interpretations of metrics inject a stream of values into the organization. Problem-solving cycles have benefited from metrics in all the phases. Metrics are used for recognition and later for diagnostics of problems. Experiments are conducted to test ideas, true to the scientific spirit of metrics application.

Software effort evaluation is one of the most essential and crucial part of software project planning for which efficient effort metrics is required. Software effort evaluation is followed by software cost evaluation which is helpful for both customers and developers. Thus, efficiency of effort component of software is very essential. Software effort evaluation is an important activity in software engineering. Estimating software effort early in software development lifecycle is a challenging task. Software size estimate is one of the most important inputs for software effort evaluation. Thus providing a size estimate with good accuracy early in the lifecycle is equally important.

However, estimates that are computed early in the lifecycle are typically associated with uncertainty. To deal with this problem, many effort evaluation techniques and metrics are developed by many researchers based on many different methods. Traditionally, there are various evaluation techniques based on comparison, analogy, equations which are broadly categorized as macro evaluation techniques. There are other micro evaluation techniques also like work breakdown structure (also known as Delphi Technique) based on Expert Judgment. The other most popular method used for effort evaluation is COCOMO found by Barry Boehm in 1981. One more technique called Putnam's Life Cycle Model is also available in the literature.

2. Background

This section describes the software effort evaluation approaches, traditional effort evaluation techniques and soft computing based evaluation for effort. It also describes the software metrics.

2.1 Effort Evaluation Approaches

There are two major software evaluation approaches: macro (for example, top-down; parametric) and micro (for example, bottom-up; task based), although some evaluation approaches combine typical aspects of both macro and micro techniques. Any of the techniques could be used at any point in the life cycle. However, the more accurate is the estimate of the project's size, the more precise is the effort and duration estimates. The relative precision of resultant estimates will match the precision of inputs.

2.2 Traditional Effort Evaluation Techniques

Based on the above mentioned approaches, there are various effort evaluation techniques in both the categories.

2.2.1 Delphi Technique

When quantified or empirical data are absent, then expertise based techniques are needed. The opinion of experts is taken, but the drawback with this technique is that the estimate is as well as the expert's opinion only. For example, Delphi technique or work breakdown structure. Delphi is a place in Greece, which was supposed to confer predictive powers to the person. A temple was built there and virgin girls were appointed there to answer questions about the future, they were called oracles. Oracle's prophecies were considered prophetic or at least wise counsel [6]. So, Delphi technique was derived from them. Under this method, project specifications are given to a few experts and their opinion is taken.

2.2.2 Putnam's Life Cycle Model

The Putnam Model is an empirical software effort evaluation model [7]. Lawrence H. Putnam in 1978 [8] is seen as pioneering work in the field of Software Process Modeling. This model describes the time and effort required for a project of specified size. SLIM (Software Lifecycle Management) is the name given by Putnam. Closely related software parametric models are COCOMO (Constructive Cost Model), PRICE-S (Parametric review of Information for Costing and Evaluation Software) and (SEER-SEM) Software Evaluation and Evaluation of Resources-Software estimating model. Nordon studied the staffing patterns of several R & D projects. He noted that the staffing pattern can be approximated by a Rayleigh distribution curve. Putnam studied the work of Nordon and determined that Rayleigh curve can be used to relate the number of lines of code to estimate time and effort required by the project.

2.2.3 COCOMO

The Constructive Cost Model (COCOMO) was launched in 1981 by Barry Boehm. It is also called COCOMO 81. The model assumes that the size of a project can be estimated in thousands of delivered source instruction and then uses a non-linear equation to determine the effort for the project. COCOMO II is the successor of COCOMO 81 and is better suited for estimating modern software development projects and updated project database. The need for the new model came as software development technology moved from mainframe and overnight batch processing to desktop development, code reusability and the use of off-the-shelf software components. COCOMO consists of a hierarchy of three increasingly detailed and accurate forms. The first level, Basic COCOMO is good for quick, early, rough order of magnitude estimates of software costs, but its accuracy is limited due to its lack of factors to account for difference in project attributes (Cost Drivers). Intermediate COCOMO takes these Cost Drivers into account and Detailed COCOMO additionally accounts for the influence of individual project phases.

2.3 Soft Computing Based Effort Evaluation Techniques

The limitations of algorithmic models led to the exploration of the non algorithmic techniques which are soft computing based. These include artificial neural network, evolutionary computation, fuzzy logic models, case-based reasoning, and combinational models and so on.

2.3.1 Neural Networks

Neural networks are nets of processing elements that are able to learn the mapping existent between input and output data. The neuron computes a weighted sum of its inputs and generates an output if the sum exceeds a certain threshold. This output then becomes an excitatory (positive) or inhibitory (negative) input to other neurons in the network. The process continues until one or more outputs are generated [9]. It reports the use of neural networks for predicting software reliability; including experiments with both feed forward and Jordan networks with a cascade correlation learning algorithm. The Neural Network is initialized with random weights and gradually learns the relationships implicit in a training data set by adjusting its weights when presented with these data. The network generates effort by propagating the initial inputs through subsequent layers of processing elements to the final output layer. Each neuron in the network computes a nonlinear function of its inputs and passes the resultant value along its output [10]. The neural network is known for its ability in tackling the classification problem. Contrarily, in effort evaluation what is needed is generalization capability.

2.3.2 Fuzzy Logic

Fuzzy logic is a valuable tool, which can be used to solve highly complex problems where a mathematical model is too difficult or impossible to create. It is also used to reduce the complexity of existing solutions as well as increase the accessibility of control theory [11]. The development of software has always been characterized by parameters that possess a certain level of fuzziness. The study showed that the fuzzy logic model has a place in software effort evaluation [12]. The application of fuzzy logic is able to overcome some of the problems which are inherent in existing effort evaluation techniques [13]. Fuzzy logic is not only useful for effort prediction, but that it is essential in order to improve the quality of current estimating models [14]. Fuzzy logic enables linguistic representation of the input and output of a model to tolerate imprecision [15]. It is particularly suitable for effort evaluation as many software attributes are measured on nominal or ordinal scale type which is a particular case of linguistic values [16].

2.3.3 Genetic Algorithm

Genetic Algorithm is one of the evolutionary methods for effort evaluation. Evolutionary computation techniques are characterized by the fact that the solution is achieved by means of a cycle of generations of candidate solutions that are pruned by the criteria ‘survival of the fittest’ [17]. When GA is used for the resolution of real-world problems, a population comprised of a random set of individuals is generated. The population is evaluated during the evolution process. For each individual a rating is given, reflecting the degree of adaptation of the individual to the environment. A percentage of the most adapted individuals are kept, while that the others are discarded. The individuals kept in the selection process can suffer modifications in their basic characteristics through a mechanism of reproduction. This mechanism is applied to the current population aiming to explore the search space and to find better solutions for the problem by means of crossover and mutation operators generating new individuals for the next generation. This process, called reproduction, is repeated until a satisfactory solution is found [18].

2.3.4 Particle Swarm Optimization

Particle swarm optimization (PSO) is a computational method that optimizes a problem by iteratively trying to improve a candidate solution with regard to a given measure of quality. Such methods are commonly known as Meta Heuristics as they make little or no assumptions about the problem being optimized and can search very large spaces of candidate solutions. PSO shares many similarities with evolutionary computation techniques such as Genetic Algorithms (GA). The system is initialized with a population of random solutions and searches for optima by updating generations. However, unlike GA, PSO has no evolution operators such as crossover and mutation. In PSO, the potential solutions, called particles, fly through the problem space by following the current optimum particles. An algorithm [19] is developed named Particle Swarm Optimization Algorithm (PSOA) to fine tune the fuzzy estimate for the development of software projects.

2.4 Software Metrics

“Metric is a quantitative measure of the degree to which a given attribute is possessed by a system or its component or by a process.” Software metrics are measures that are used to quantify different attributes of a software product, software development resource and software development process. *Software metrics* deals with the evaluation and measurement of different attributes of the software product and the software development process. [1]

3. Literature Review

This section explores some of the researches done on software evaluation using different techniques by various researchers in previous years.

In **2010**, **Jin-Cherng Lin** et al. [20] used Pearson product moment correlation coefficient and one-way to analyze to select several factors and then used K-Means clustering algorithm to software project clustering. After project clustering, they used Particle Swarm Optimization that takes the mean of MRE (MMRE) as a fitness value and N-1 test method for optimization of COCOMO parameters. Finally, they took parameters that finish the optimization to calculate the software project effort that is wanting to evaluate. This research used 63 history software project data of COCOMO to test. The experiment really expresses using base on project clustering with multiple factors making more effective base on the effort of the software estimate of COCOMO's three project mode.

In **2011**, **Jin-Cherng** Lin et al. [21] proposed a model which combines genetic algorithm (GA) with support vector machines (SVM). We can find the best parameter of SVM regression of the proposed model, and make more accurate predictions. The model was tested and verified by using the historical data in COCOMO, Desharnais, Kemerer, and Albrecht. The results were shown by prediction level (PRED) and the mean magnitude of relative error (MMRE).

In **2012**, **Thamarai.I** et al. [22] proposed a genetic algorithm and artificial neural network based on which Feature Selection and Similarity Measure between the projects can be achieved by using Differential Evolution. This is a population based search strategy. The Differential Evolution is used to compare the key attributes between the two projects. Thus we can get most optimal projects which can be used for the evaluation of effort using the analogy method.

4. About the Problem

To get an accurate or near to accurate effort evaluation has always been a challenge in software development. To deal with this problem, many researchers have contributed in various areas by applying many techniques. These techniques include regression analysis, analogy-based evaluation, comparison based evaluation and machine learning based evaluation. The uncertainty can be reduced by any of the above techniques.

Neural networks are good at training the dataset, but the clustering and feature input is somewhat weak in neural networks. Fuzzy logic based models are good at featuring and clustering but the training of datasets is not provided. Genetic algorithms as an optimization technique are usually applied in multi-neural systems in order to improve operations or performance of the system, either as an expert or global level. PSO easily suffers from the partial optimism, which causes the less exact at the regulation of its speed and the direction. The method cannot work out the problems of scattering and optimization. The method cannot work out the problems with non-coordinate system, such as the solution to the energy field and the moving rules of the particles in the energy field.

From the above survey, it is clear that each of the methods mentioned has some of the disadvantages over the other. To overcome this problem, this paper gives the detail of an efficient framework for effort evaluation using neuro-fuzzy technique i.e., ANFIS.

5. Proposed Framework

The proposed framework includes the evaluation of software effort using neuro-fuzzy based (ANFIS toolbox) of MATLAB. The details for proposed framework are mentioned as under:

- For software effort evaluation, NASA dataset with 18 projects is considered for implementation. The performance measures MMRE and RMSE are used for comparing the performance of ANFIS in effort evaluation with other traditional evaluation models.

Thus, the framework measures the effort component of the software efficiently using ANFIS which in turn is useful for cost evaluation of software.

This paper proposes ANFIS based software effort evaluation. ANFIS is a hybrid AI technique, which combines best features of Fuzzy Logic and parallel processing neural networks. It possesses fast convergence and has more accuracy than back propagation neural network. Various forms of ANFIS methods are explored for effort evaluation. ANFIS methods are comparatively good at evaluation than complex neural networks.

The Sugeno based Fuzzy Inference system is developed and in order to train the Sugeno FIS, Adaptive Neuro-Fuzzy system (ANFIS) is designed that makes use of the Sugeno FIS Structure as shown in Fig. 1.

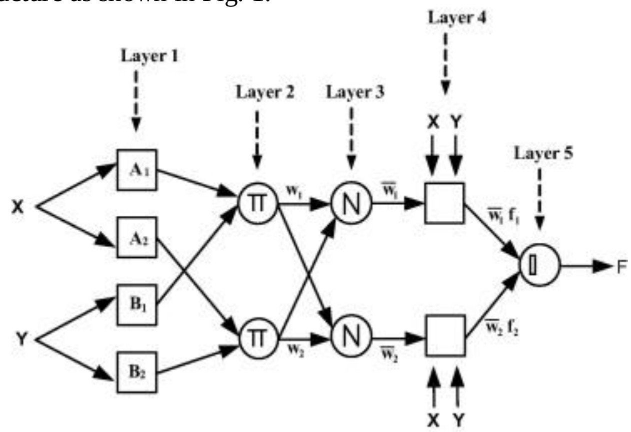


Fig.1. Architecture of ANFIS

The algorithm used for effort evaluation is based on Neuro-Fuzzy technique. More specifically, it is known as Adaptive Neuro-Fuzzy Inference System (ANFIS). The implementation is done on a NASA dataset of 18 projects in MATLAB R2011a Environment. The steps of the proposed algorithm are shown in the form of the flowchart in Figure 2.

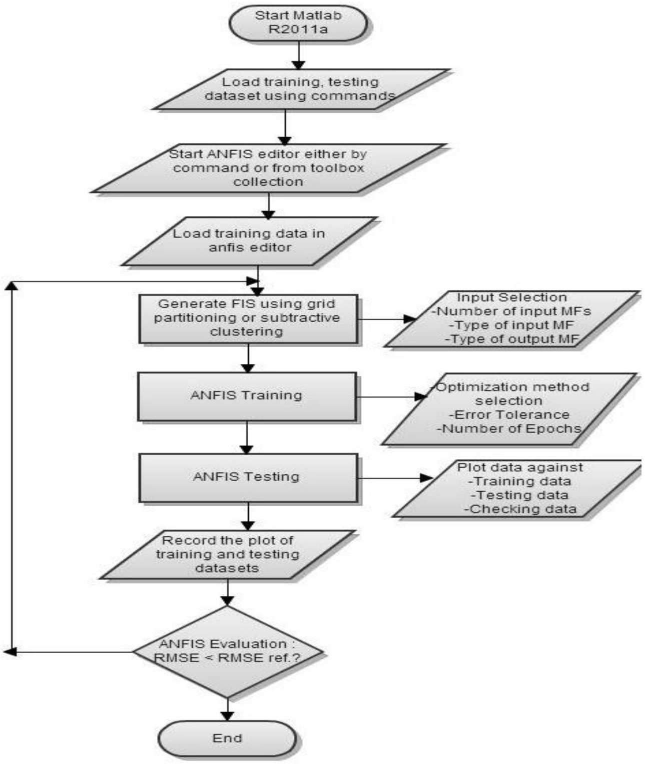


Fig.2. Flowchart of ANFIS

6. Experimentation and Results

The list of parameters used for simulation in MATLAB is shown in the Table 1 below:

Table 1: List Of Parameter Variables And Their Values

Parameter Variables	**Associated Values**
Simulation Tool	MATLAB 7.12.0 (R2011a)
Dataset used for experimentation	NASA dataset
Total No. of projects	18
No. of projects used for training	13
No. of projects used for testing	5
FIS method	Grid Partitioning
Optimization method	Hybrid
No. of membership functions	2
Type of membership functions	Trimf, Trapmf, gbellmf, gaussmf, gauss2mf, pimf, dsigmf, psigmf.
No. of epochs	500

In Table 2 and 3 the computed effort for training datasets is described for each membership function using ANFIS toolbox of MATLAB. In Table 4 and 5 *RMSE* and *MMRE* criteria is computed over the complete data set for ANFIS model for different membership functions is shown [23]. In Table 6 computed software quality metric for membership functions compared with the other models are shown.

Neuro-Fuzzy model using ANFIS toolbox of MATLAB uses different membership functions. There are 8 functions in ANFIS, out of which gauss2 membership function has the lowest MMRE and RMSE of 0.0050 and 0.6410 respectively. Also, gbell membership function has the lowest MMRE and RMSE of 0.0367 and 0.4976.

The software quality metric EEA should approach to 1 and SP (total) should be equal to the value resulting from total source size divided by the actual effort value. Here, in this case, we have taken the comparison for a first project from datasets of all projects. Accordingly, the value of EEA and SP (total) for neuro-fuzzy functions outperforms the other traditional models.

Table 2: Computed Effort For Nasa Software Projects-Training Case - ANFIS Functions

No.	DKLOC	Method-ology	Actual Effort	Trimf Effort	Trapmf Effort	Gbellmf Effort	Gaussmf Effort
1	90.2	30	115.8	115.7801	115.7990	115.8005	115.7801
2	46.2	20	96	95.6446	88.6000	95.9679	95.6446
3	46.5	19	79	79.2911	88.6000	79.0235	79.2911
4	54.5	20	90.8	90.7932	88.6000	90.8010	90.7932
5	31.1	35	39.6	39.8002	39.6113	39.6464	39.8002
6	67.5	29	98.4	98.4448	98.4015	98.3999	98.4448
7	12.8	26	18.9	19.9137	10.2485	18.9679	19.9137
8	10.5	34	10.3	9.4786	8.8195	9.2436	9.4786
9	21.5	31	28.5	27.9643	28.4680	28.4765	27.9643
10	3.1	26	7	5.7052	10.2485	6.2696	5.7052
11	4.2	19	9	9.0087	9.0000	9.0504	9.0087
12	7.8	31	7.3	8.4028	8.9545	8.4367	8.4028
13	2.1	28	5	5.3722	10.2485	5.5157	5.3722

Table 3: Computed Effort For Nasa Software Projects-Training Case - ANFIS Functions

No.	DKLOC	Method-ology	Actual Effort	Gauss2mf Effort	Pimf Effort	Dsigmf Effort	Psigmf Effort
1	90.2	30	115.8	115.7970	115.7995	115.7972	115.7972
2	46.2	20	96	95.9509	88.6000	95.9353	95.9334
3	46.5	19	79	79.0113	88.6000	79.0390	79.0396
4	54.5	20	90.8	90.8375	88.6000	90.8235	90.8247
5	31.1	35	39.6	39.6263	39.6083	39.6210	39.6154
6	67.5	29	98.4	98.4053	98.4008	98.4048	98.4048
7	12.8	26	18.9	18.8415	10.2625	18.8965	18.8706
8	10.5	34	10.3	9.0406	8.8123	9.0812	9.0714
9	21.5	31	28.5	28.4568	28.4770	28.4656	28.4914
10	3.1	26	7	6.0100	10.2625	6.0194	6.0301
11	4.2	19	9	9.0034	9.0000	9.0039	9.0040
12	7.8	31	7.3	8.6562	8.9143	8.7286	8.7210
13	2.1	28	5	5.9624	10.2625	5.7833	5.7953

Table 4: Computed RMSE And MMRE Criterion For Anfis Functions

Performance Criteria	**ANFIS Function Used**			
	trimf	**trapmf**	**gbellmf**	**gaussmf**
MMRE	0.1770	0.1974	0.0367	0.0443
RMSE	0.6369	4.5543	0.4976	0.6369

Table.5 Computed RMSE And MMRE Criterion For Anfis Functions

Performance Criteria	**ANFIS Function Used**			
	gauss2mf	**pimf**	**dsigmf**	**psigmf**
MMRE	0.0050	0.1973	0.0472	0.0473
RMSE	0.6410	4.5533	0.6269	0.6269

Table 6: Computed EEA and SP (total) Criterion for ANFIS Functions for project 1 in the dataset

Software quality metric	trimf	trapmf	gbellmf	gaussmf	gauss2mf	pimf	dsigmf	psigmf
EEA	1	1	0.9999	1.0001	1	1	1	1
SP (total)	0.799	0.7789	0.7789	0.779	0.7789	0.779	0.7789	0.7789

7. Conclusion and Future Work

This paper evaluates software effort efficiently using ANFIS based learning techniques. Accurate evaluation of effort leads to other evaluations efficiently and accurately like cost, staffing, budget and schedule. Effort component of software plays a vital role in software project management. By predicting software effort, proposed ANFIS based technique may facilitate the software planning stage in making its decision regarding the evaluation of other resources of the software. The ANFIS based technique was successfully implemented to predict software effort. The same was compared with neural network based technique and other models which were previously reported. In the basic scheme three types of ANFIS were used for learning. All three provided better performance in all performances metric with respect to the neural network. Conventional ANFIS worked better in accuracy and RMSE error compared all type neural networks and other previous methods.

Software effort doesn't only depend on the size of the project, it may include different other parameters like Intermediate COCOMO attributes. So ANFIS based technique must be tuned to predict all these attributes affecting software effort. So for these different types of attributes hybrid ANFIS must be explored. In the extended case only the methodology and size of the project parameters are included but practical situations also affect effort and other resources. The problem must be formalized to include other parameters which affect effort. The prediction was based on an assumed scenario but to validate and check the robustness of ANFIS more realistic time series must be considered for training.

1. Analysis can be made for another type of datasets like ISBSG, IBM etc.
2. Calculation of Pred (25), Spearman's rank can lead to better validation of prediction models.
3. Analysis can also be done using artificially generated data set.
4. Analyzing the performance of the model by varying the number of epoch, number of membership functions.

References

[1] Jibitesh Mishra, Ashok Mohanty, (2011) "Software Engineering", CH-13, Software Metrics, Pearson Education India.

[2] Tu Honglei, Sun Wei, and Zhang Yanan, "The research on software metrics and software complexity metrics", in Computer Science-Technology and Applications, 2009. IFCSTA '09. International Forum, volume 1, pages 131 –136, Dec. 2009.

[3] S.R. Chidamber and C.F. Kemerer. "A metrics suite for object oriented design", IEEE Transactions on Software Engineering, 20(6):476 –493, Jun. 1994.

[4] M. D'Ambros, M. Lanza, and R. Robbes. An extensive comparison of bug prediction approaches. In 7th IEEE Working Conference on Mining Software Repositories (MSR), 2010, pages 31 –41, May 2010.

[5] V.R. Basili, L.C. Briand, and W.L. Melo. "A validation of object-oriented design metrics as quality indicators", IEEE Transactions on Software Engineering, 22(10):751 –761, Oct. 1996.

[6] Murali Chemuturi, Delphi Technique for software evaluation.

[7] Putnam, Lawrence H.; Ware Myers (2003). Five core metrics: the intelligence behind successful software management. Dorset House Publishing. ISBN 0-932633-55-2.

[8] Putnam, Lawrence H. (1978). "A General Empirical Solution to the Macro Software Sizing and Estimating Problem". IEEE transactions on Software Engineering, VOL. SE-4, NO. 4, pp 345-361.

[9] N. Karunanitthi, D. Whitley, and Y. K. Malaiya, (1992), "Using Neural Networks in Reliability Prediction", IEEE Software, Vol. 9, no.4, pp. 53-59.

[10] Idri, T. M. Khoshgoftaar, A. Abran, (2002), "Can neural networks be easily interpreted in software cost evaluation?" IEEE Trans. Software Engineering, Vol. 2, pp. 1162 – 1167.

[11] Razaz, M. and King, J. (2004)"Introduction to Fuzzy Logic" Information Systems - Signal and Image Processing Group. http://www.sys.uea.ac.uk/king/restricted/boards/.

[12] Moon Ting Su1, Teck Chaw Ling, Keat Keong Phang, Chee Sun Liew, Peck Yen Man, (2007), "Enhanced Software Development Effort And Cost Evaluation Using Fuzzy Logic Model", Malaysian Journal of Computer Science, Vol. 20(2), pp 199-207.

[13] R. Gray, S. G. MacDonell, (1997), "Applications of Fuzzy Logic to Software Metric Models for Development Effort Evaluation", Fuzzy Information Processing Society 1997 NAFIPS' 97, Annual Meeting of the North American 21 - 14 pp. 394-399.

[14] S. Kumar, B. A. Krishna, and P. S. Satsangi, (1994), "Fuzzy systems and neural networks" in software engineering project management, Journal of Applied Intelligence, no. 4, pp. 31-52.

[15] N. E. Fenton, S. L. Pfleeger, (1997),"Software Metrics, a Rigorous and Practical Approach", 2nd Edition, PWS Publishing Company, Thomson Publishing,Boston.

[16] Agustin Gutierrez T., Cornelio Yanez M. and Jerome Leboeuf Pasquier, (2005), "Software Development Effort Evaluation Using Fuzzy Logic: A Case Study", Proceedings of the Sixth Mexican International Conference on Computer Science (ENC'05).

[17] Urkola Leire , Dolado J. Javier , Fernandez Luis and Otero M. Carmen , (2002), "Software Effort Evaluation: the Elusive Goal in Project Management", International Conference on Enterprise Information Systems, pp.412-418.

[18] A.P. Engelbrecht, (2006), Fundamentals of Computational Swarm Intelligence, John Wiley & Sons, New Jersey.

[19] Prasad Reddy, (2010), "Particle Swarm Optimization in the fine-tuning of Fuzzy Software Cost Evaluation Models, International Journal of Software Engineering (IJSE), Volume (1): Issue (1), pp 12-23.

[20] Jin-Cherng Lin, Han-Yuan Tzeng (2010), "Applying Particle Swarm Optimization to Estimate Software Effort by Multiple Factors Software Project Clustering", Computer Symposium (ICS), Proceedings published in IEEE, pp. 1039-1044.

[21] Jin-Cherng Lin, Chu-Ting Chang and Sheng-Yu Huang (2011), "Research on Software Effort Evaluation Combined with Genetic Algorithm and Support Vector Regression", International Symposium on Computer Science and Society, Proceedings published in IEEE, pp. 349-352.

[22] Thamarai.I, Dr.S.Murugavalli (December 2012), "Using Differential Evolution in the Prediction of Software Effort", IEEE Fourth International Conference on Advanced Computing (ICoAC), pp. 1-3.

[23] http://www.mathworks.in/help/fuzzy/anfis-and-the-anfis-editor-gui.html.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 1, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

Metric Base Analysis and Modeling Experiments of Routing Protocols in MANETs and VANETs Wireless Network using Real Time Scenarios

Syed Fakhar Abbas[1], Ghulam Yasin[2],Muhammad Akram Mujahid[3],Dr. S R Chaudhry[4]

[1] Communication Network Research Center (CNRC) COMSATS Institute of Information Technology
Lahore, Pakistan

[2] Divisions of Science and Technology, University of Education
Lahore, Pakistan

[3] Divisions of Science and Technology University of Education
Lahore, Pakistan

[4] Communication Network Research Center (CNRC) COMSATS Institute of Information Technology
Lahore, Pakistan

Abstract

Ad hoc network is a collection of wireless mobile hosts forming a temporary network without the aid of any established infrastructure or centralized administration. In such an environment, it may be necessary for one mobile host to enlist the aid of other hosts in forwarding a packet to its destination, due to the limited range of each mobile host's wireless transmissions. Mobile ad hoc networks (MANET) do not rely on any fixed infrastructure but communicate in a self-organized way. Vehicular Ad hoc Network (VANET) is a subordinate of Mobile Ad hoc Network (MANET). In this paper, we have discussed experimental metrics base framework in which Data Drop (DD) and throughput are measured and analyzed for Mobile Ad-hoc Network (MANETs) and Vehicular Ad-hoc Network (VANETs). Moreover, the authors have contributed the performance comparison of routing protocols. Novel contribution in this paper is to compare different routing protocols in both MANETs and VANETs. The proposed system classifies the changing in MANET nodes and VANET vehicles data drop, and throughput. It was observed from the results that in form of high through put and low packet drop DSR shows better performance compared to DSDV and AODV in both VANET and MANET. Protocols which are being analyzed in this paper are: AODV, DSR and DSDV. The performance parameter includes Data Dropped and throughput.

Keywords: *VANET, MANET, AODV, DSR, OLSR, DSDV*

1. INTRODUCTION

Wireless technology in the field of has been growing fast previous more than two decades. Developments have played a vital role to explore new horizons doors for researchers to dig out unique cost-effective solutions for different applications. Cellular networks, Ad hoc networks, ,wireless sensor networks, Visible Light Communication, Wi-Fi, Wi-Max are some of the illustrations of new wireless network technologies that have been used in security ,telecommunication, engineering, location tracking , network monitoring , remote sensing, , medical, education, and tracking systems.

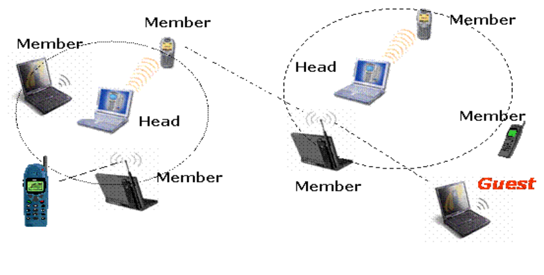


Fig. 1. Communication architecture of MANET

The proliferation of wireless communication and mobile devices in recent years has opened the door to research on self-organizing networks that do not require a pre-established infrastructure as shown in Fig. 1. Those spontaneous networks, normally called ad hoc networks, provide mobile users with ubiquitous communication capacity and information access regardless of the location.

The most important characteristic of such network is the independence of any fixed infrastructure or centralized administration. An ad hoc network is capable of operating autonomously and is completely self-organizing and self-configuring. Therefore, it can be rapid and easily deployed.

Another important property of an ad hoc network is the multihop capability. Unlike the cellular networks, which are single-hop wireless networks, an ad hoc network does not guarantee that a mobile node can directly communicate with its destinations all the time. A mobile node, which lies outside the transmission range of its specific destination, would need to relay its information flow through other mobile nodes. This implies that mobile nodes in ad hoc networks bear routing functionality so that they can act both as routers and hosts.

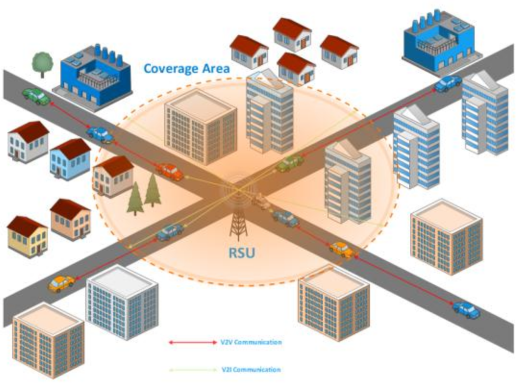


Fig. 2. Communication architecture of VANET

Fig. 2 shows the architecture of VANET in which different components are shown. This is generic architecture in which data is transferred among different on-board units (OBU). Intelligent Transportation Systems (ITS) is a wide-ranging technology system applicable to transportation to make system safer, more effective, and more reliable and more environment friendly, without altering the existing infrastructure. Technologies ranges include sensor network, control technologies, communications system, computer informatics, transportation, engineering, telecommunications, computer science, finance, electronic commerce and automobile manufacturing. A vehicular ad hoc network (VANET) is an emerging research area for the communications industry and academician. Researchers proposed an entirely new wireless networking concept i.e. vehicular ad hoc network which can increase passenger safety and reduces vehicle collisions on the road. VANET can provide an efficient road, comfort facility to the passengers and traffic regulations monitoring in future. Wireless communication among moving vehicles is unique and innovative research era in the academics and in the corporate sector, driven by the vision to communicate information among vehicles to ensure the safety and comfort of the users [1–2]. Now a days automobile industry have equipped their new vehicles with global positioning systems (GPS), digital maps and even wireless interfaces, e.g. Honda-ASV3.The network architecture of VANET can be classified into three categories: Cellular network/WLAN, ad hoc network, and hybrid network [3].

This paper summarizes the impact of topology based routing protocols in MANET and VANET wireless scenarios. The remainder of the paper is organized as follows: Section II describes the problem statement. Section III discusses the different routing protocols used in VANET and MANET. Section IV presents the experimental and model scenarios & V analyze the results of urban and rural area scenarios. We conclude in Section VI and section VII for reference.

2. PROBLEM STATEMENT

With the emergence of vigorous wireless network technologies and wireless applications intensive research have been conducted in VANET and MANET. Vehicles are needed to equip with Global Positioning System (GPS) and IEEE 802.11 wireless adaptors to create ad-hoc network for data sharing. Ad-hoc routing protocols are susceptible to move during data transmission in the network. Path is pre-established for data transfer and network configuration changes frequently.

Challenges faced by ad-hoc network includes security, bandwidth limitation, energy utilization, scalability, network performance etc. while measuring different parameters mentioned above we use different protocols to show the network performance and quality of service. The proposed system classifies the changing vehicular density, data drop, and throughput and end-to-end delay of nodes and vehicles in the network.

To guarantee the performance of network multiple protocols have been analyzed in the simulation scenario like AODV, DSR and DSDV. Performance has been measured in rural and urban areas using parameters like: data drop, throughput and end-to-end delay of nodes in both VANET and MANET network.

3. TOPOLOGY BASED ROUTING PROTOCOLS

In this paper, we have only considered only proactive and reactive protocols. Based routing protocols usage is link's information, which exists in the network and use as packet forwarding. These protocols can be categorized into:

- Proactive (table-driven) routing protocols
- Reactive (on-demand) routing protocols
- Hybrid routing protocols

3.1 Proactive-Routing Protocols

Proactive protocols based mostly on algorithm's shortest path. Information is kept of all nodes connected in table form and are also shared with their neighbors [4]. Even if the paths are not being used currently, they do maintain and give updates on routing amongst all the vehicles for given network at any times, therefore even in cases where paths are never used, updates regarding them are continuously broadcasted among nodes [5]. When periodical route updates are performed, network load, constraints of bandwidth and size of the network are never regarded. This is a major limitation of using this approach in VANETs. Types of Pro-active Routing Protocols: FSR, DSDV, PLSR, CGSR, WRP and TBRPF.

1) Destination Sequenced Distance Vector (DSDV)

It is Table Driven routing protocol which is used in VANET and is grounded on classical Bellman-Ford algorithm. Primarily every vehicle broadcasts its own route information tables to its neighbor vehicles. The neighbor vehicles keep up-to-date routing table by two type of packets- Full Dump Packet and Incremental Packet. Full Dump Packet comprises information about every contributing vehicle in the VANET. These packets are communicated intermittently after a long time intermission. Incremental Packet covers latest change in vehicle position since last Full Dump Packet. These packets are communicated periodically in short interval of time and their information is stored in additional table. Routes are nominated with the up-to-date entry in the table. DSDV is better option for networks where location of nodes is less changeable. If position of a vehicle changes frequently, its performance will decreases due to more Full Dump Packets are required to send in the network, which results the bandwidth wastage.

3.2 Reactive Routing Protocols

Reactive routing protocols for mobile ad hoc networks are similarly known as "on-demand" routing protocols. In a reactive destination node but no route is accessible, it initiates a route detection process. It is initiated with RREQ packet, response is with RREP and while link is not available it is received RERR packet. Reactive routing protocols has less overhead, a unique feature, while reactive routing protocols scalable than proactive routing protocols. But while using reactive routing protocols, source nodes may undergo routing protocol, routing paths are look for only while it is desirable. Fig. 3 shows the classification of AODV showing the procedure of RREQ and RREP. Fig. 4 shows the classification of DSR showing the procedure of RREQ and RREP. Henceforth these protocols are not appropriate for real–time applications. The Dynamic Source Routing (DSR) [6] and Ad hoc On-demand Distance Vector routing (AODV) [6] are examples for reactive routing protocols.

2) Ad Hoc on-Demand Distance Vector Routing (AODV)

AODV protocol is an enhancement on DSDV protocol. When two nodes require a connection, AODV maintains overall routing tables in the nodes and establishes routes only. AODV classification is as a pure on demand route acquisition system [7]. This is because a small number of routing information is kept in the nodes which are not on the route selected.

A distinguishing feature of AODV is that it tries to avoid the "counting to infinity" problem by using sequence number in routing discovery process [8]. Three types of messages used in AODV routing process are Route Error (RERR), Route Request (RREQ) and Route Reply (RREP).The network maintains silence until when the source nodes requires communicating with the node of communication. In case of lack of valid route that existed, an initiation of route discovery process will begin. The source of the node then broadcasts RREQs to its neighbors, then RREQ is forwarded to their neighbors and it continues on and on until a point is reached which is the RREQs destination or an intermediate node that is the destination's valid route. A sequence number and broadcast ID will be assigned by each node. An increase in the sequence number is done along the path with every generation of RREQ. Consequently, identification of RREQ can be done based on the IP address and broadcast ID based on the source node.

RREQ replies back to the source when it reaches the destination or an intermediate node that has a route that is active. All the other RREQ that arrives later from other nodes are done away with. Link failure detection by upstream node within the network makes RERR to be fed back to the source node and an initiation of new route discovery process begins after receiving RERR. Fig. 3 shows the classification of AODV showing the procedure of RREQ and RREP.

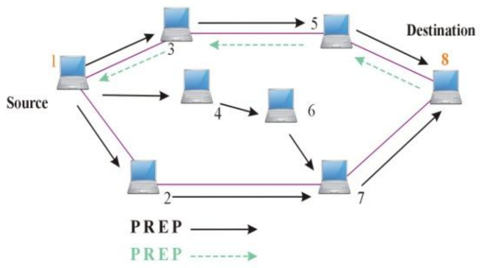


Fig. 3. AODV Protocol Behavior

3) Dynamic Source Routing (DSR)

DSR is based on the source concept routing [9] it is uncomplicated ad hoc routing protocol. It is Source Initiated on Demand routing protocol used in VANET and is grounded on link state routing algorithm. When a vehicle needs to transfer data to another vehicle, first it initiates route discoveries request up to that vehicle. For route finding, source vehicle recruits a route request (RREQ) packet in the network and other nodes forward the RREQ by changing their name as sender. Lastly when RREQ packet spreads to the destination vehicle or to a vehicle having path to the destination vehicle, a route reply (RREP) packet is unicasted to the sender node. If the reply packet is not received, the source vehicle resumes violent discovery of route up to the destination vehicle. Fig. 4 shows the classification of DSR showing the procedure of RREQ and RREP.

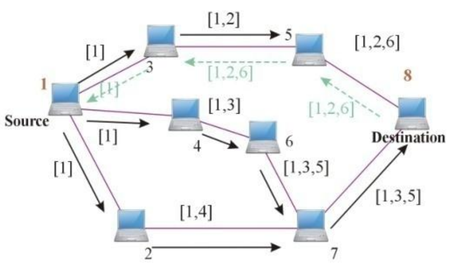


Fig. 4. DSR Protocol Behavior

Table I Comparisons of routing protocols

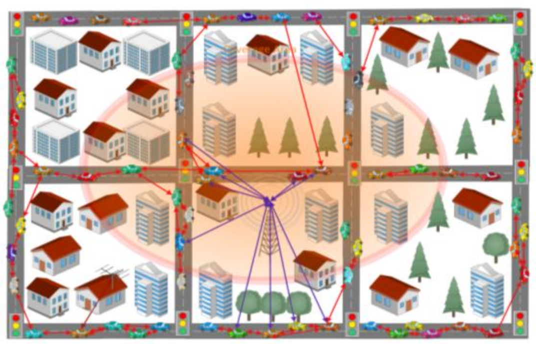
Protocol Property	DSDV	DSR	AODV
Loop Free	Yes	Yes	Yes
Multicast Routes	No	Yes	No
Distributed	Yes	Yes	Yes
Unidirectional Link Support	No	Yes	No
Multicast	No	No	Yes
Periodic Broadcast	Yes	No	Yes
Routes Preserved in	Route Table	Route Cache	Route Table
Route Cache	Yes	No	Yes
Reactive	No	Yes	Yes

4. EXPERIMENTAL MODEL AND SCENARIO

In the MANET and VANET networks basically NCTUns 6.0 is used which is the free version of NCTUns network simulator and emulator, which is a world-renowned tool and has been used by more than 20,000 listed users from 144 countries all over the world. EstiNet 7.0 is latest version and whose central technology is based on the novel kernel re-entering methodology invented by Prof. S.Y. Wang [10]. The various features of VANET supported by NCTUns 6.0 make it a clear choice for proposed research work.

4.1 VANET Urban Area Scenario

Fig. 5 shows the urban area grid scenario where 80 vehicles devices identified as On-Board Units (OBUs) communicate with each other as well as with RSU (Road Side Unit) .Vehicles show the network behavior as the OBUs move within the network to analyze the performance of each protocol. While assessing the performance of a given scenario in the vehicle-to vehicle communication (V2V) and vehicle-to-roadside communication (V2R) vehicles move within network and establish VANET. In this mobility model we used Random waypoint. Using this mobility model vehicles



are free to move to reach at random destination. Movement of the vehicles is calculated by the algorithm.

Fig. 5. VANETs Urban Area Grid Scenario

4.2 MANET Scenario

Fig. 6 shows MANET scenario where 80 MANET nodes communicate with each other.

Fig. 6. MANET Scenario

5. Performance Results of AODV, DSR, DSDV

Following graphs show the performance of the routing protocol using different metric considered above. The X- Axis demonstrations the time in minutes and the y axis displays the Metric considered.

5.1 Throughput

From fig. 7 it can be realized that in VANET area scenario throughput with DSR protocol is better than with AODV and DSDV with throughput peak reaching up to 300 kB/s. It is also observed that speed is increased and more vehicles connected to the RSU than AODV performance suddenly degrades from about 270 KB/s to 40 KB/s while DSDV performance slightly decrease remains moderately same.

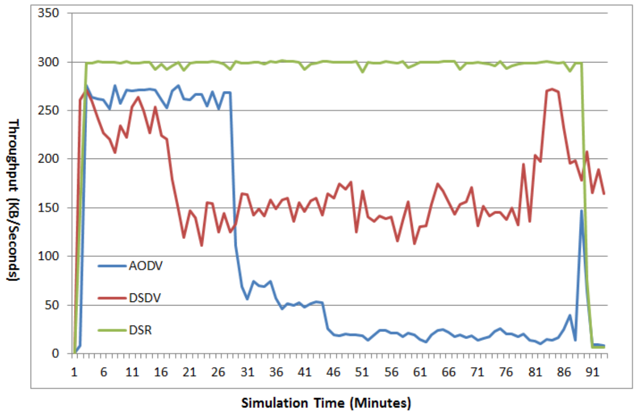


Fig. 7. Throughput of VANET Scenario

Fig. 8 shows the Throughput in MANET scenario is exposed. Clear implication from graph DSR Throughput is uppermost. AODV Throughput remains in between other two and DSDV Throughput is lowest among all three.

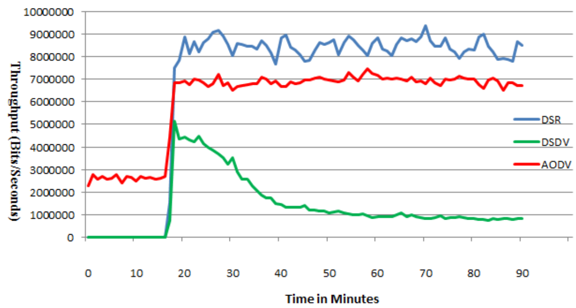


Fig. 8. Throughput of MANET Scenario

5.2 Packet Dropped

The packet drop performance of DSR protocol is much better than AODV and DSVDV in VANET scenario as is seen from fig. 9. About 10 to 20 packets dropped in urban area while less than 50 packets drop in rural scenario while this drop of packets better as time passes for the DSR protocol. As the speed is increased the packet drop rate for AODV protocol increases from 200 to 1100 drop packets in urban area whereas in rural area packet drop for AODV remain between 100 to 140 packets. In terms of dropped packets for DSDV's performance is the worst in both scenarios. The performance degrades with the increase in the number of nodes and speed.

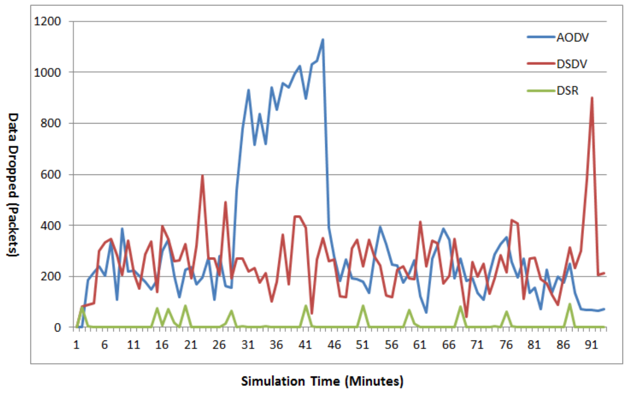


Fig. 9. Packets Dropped in VANET Scenario

Figure 10 shows the drop packet performance of DSR protocol is better than for both AODV and DSDV in MANET scenario.

As the number of vehicles increased data dropped ratio of DSDV suddenly increased.

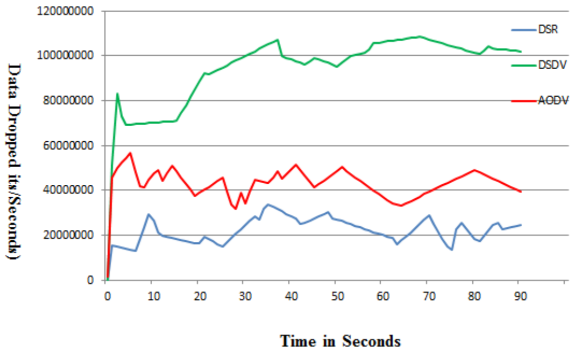


Fig. 10. Packets Dropped in VANET Scenario

6. CONCLUSION

In this paper, presented simulation studies and compared the Routing protocol DSDV uses proactive “table driven” routing, while AODV and DSR use reactive “on-demand” routing. Protocol DSDV periodically updates its routing tables, even in cases when network topology doesn’t change. AODV protocol has inefficient route maintenance, because it has to initiate a route discovery process every time network topology changes. Both protocols, AODV and DSR, use route discovery process, but with different routing mechanisms. In particular, AODV uses routing tables, one route per destination, and destination sequence numbers as a mechanism for determining freshness of routes and route loops prevention. On the other hand, DSR uses source routing and route caching, and doesn’t depend on any periodic or time-based operations. From the parameter values characterizing the two traffic scenarios, DSR is found suitable for both rural and urban traffics scenarios. Thus it can be concluded that DSR outperforms from other routing protocols AODV and DSDV in both urban and rural area scenarios. Poor performances of DSR routing protocol, when mobility or load are increased, are the consequence of aggressive use of caching and lack of any mechanism to expire stale routes or determine the freshness of routes when multiple choices are available.

7. REFERENCES

[1] Jedrzej Rybicki, Bj¨orn Scheuermann, Wolfgang Kiess, Christian Lochert, PezhmanFallahi,and Martin Mauve, "Challenge: peers on wheels - a road to new traffic information systems," In Proceedings of the 13th Annual ACM International Conference on Mobile Computing and Networking 2007, p 215–221.

[2] Jakob Eriksson, Hari Balakrishnan, and Samuel Madden, "Cabernet: Vehicular content delivery using WiFi" In Proceedings of the 14th ACM International Conference on Mobile Computing and Networking 2008, p 199–210.

[3] Kevin C. Lee, Uichin Lee, Mario Gerla, "Survey of Routing Protocols in Vehicular Ad Hoc Networks," Advances in Vehicular Ad-Hoc Networks: Developments and Challenges, IGI Global, Oct, 2009.

[4] Survey of Routing Protocols in Vehicular Ad Hoc Networks," Kevin C. Lee, Uichin Lee, Mario Gerla, Advances in Vehicular Ad-Hoc Networks: Developments and Challenges, IGI Global, Oct, 2009.

[5] IEEE Draft P1609.0/D01, February 2007.

[6] Hadi Sargolzaey, Ayyoub Akbari Moghanjoughi and Sabira Khatun, "A Review and Comparison of Reliable Unicast Routing Protocols For Mobile Ad Hoc Networks," IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.1, pp. 186-196, January 2009.

[7] C. E. Perkins & E. M. Royer, "Ad-hoc On-Demand Distance Vector Routing," Proc. 2nd IEEE Wksp. Mobile Comp. Sys. And Apps, Feb. 1999, pp. 90–100.

[8] Ashwin Perti, Pradeep Sharma, "Reliable AODV Protocol for Wireless Ad Hoc Networking," IEEE International Advance Computing Conference 2009.

[9] Martinez-Alfaro, H., Hernandez-Vazquez, M.A.,"DSR protocol scalability optimization in mobile ad-hoc networks with ANFIS ," Systems, Man and Cybernetics, IEEE International Conference on Vol.4 pp.3414 - 3419, 5-8 October. 2003.

[10] Shie-Yuan Wang; Chih-Che Lin, "NCTUns 5.0: A Network Simulator for IEEE 802.11(p) and 1609 Wireless Vehicular Network Researches," Vehicular Technology Conference, 2008. VTC 2008-Fall. IEEE 68th, pp.1-2, 21-24 September.

Mapping Wireless Sensor Network Applications Requirements to Existing Operating Systems

Tarek M.Salem[1], Sherine M. Abd El-kader[2]

[1] **Assistant Research at Electronics Research Institute, Computers and Systems Dept, Cairo, Egypt**

[2] **Associate Professor at Electronics Research Institute, Computers and Systems Dept, Cairo, Egypt**

Abstract

The design of operating system for Wireless Sensor Network (WSN) deviates from traditional operating system design due to significant and specific characteristics like constrained resources, high dynamics and inaccessible deployment. The purpose of this work is to classify existing operating systems according to the important Operating System (OS) features and to propose the suitable OSs for different categories of WSN applications. Architecture, execution model, scheduling, routing protocols, hardware support, and application support are the important OS features that are chosen to classify the existing WSN operating systems. This classification helps in understanding the contrasting differences between the existing operating systems and lays the foundation for designing an ideal operating system. To help the application developer in choosing the right OS, based on the application requirement, hardware type, also WSN applications have been classified. This classification gives insight in choosing the best suitable operating systems that fits for different categories of applications.

Keywords: *Wireless sensor network, Operating systems, Embedded operating system, Real-time operating system, Application requirements.*

1. Introduction

A wireless sensor node is a good example for a System on Chip (SoC) that has communication, computation, sensing and storage capabilities. These miniaturized nodes have stringent constraints in terms of available resources like processing power, battery power, program memory, available bandwidth. Basically, each node comprises of a micro-controller, power source, Radio Frequency (RF) transceiver, external memory, and sensors. These sensor nodes collectively form a Wireless Sensor Network, which is used in wide variety of applications now days [1] [2]. A WSN typically consists of hundreds or thousands of sensor nodes. These nodes have the capability to communicate with each other using multi-hop communication. Typical applications of these WSN include but not limited to monitoring, tracking, and controlling.

The basic functionality of an operating system is to hide the low-level details of the sensor node by providing a clear interface to the external world. Processor management, memory management, device management, scheduling policies, multi-threading, and multitasking are some of the low level services to be provided by an operating system. In addition to the services mentioned above, the operating system should also provide services like support for dynamic loading and unloading of modules, providing proper concurrency mechanisms, Application Programming Interface (API) to access underlying hardware, and enforce proper power management policies.

Though some of these are similar to the services provided by traditional operating systems, the realization of those services in WSN is a non-trivial problem, due to the constraints on the resource capabilities. Hence a suitable operating system is required for WSN to provide these functionalities to facilitate the user in writing applications easily with little knowledge of the low-level hardware details.

Due to the significance of an operating system for WSNs and the availability of a significant body of literature on it, study of search becomes necessary and useful at this stage. Also analyzing different applications, their characteristics and suggesting an ideal OS for those applications will help an application developer to choose an operating system and therefore select best sensor hardware for specific application. Although there are many papers that surveys the characteristics, applications, and communication protocols of WSNs [3], prior to this there are no studies that survey the operating systems of WSN for specific application.

The rest of this paper is organized as follows. In Section 2, describes related work for this work. Section 3 describes design requirements of an operating system for WSNS. A classification framework and a comprehensive survey of existing operating systems against this framework are presented in Sections 4, 5 respectively. Comparative analysis between different operating systems is tabulated in section 6. Conclusions and future work are introduced in Section 6, 7 respectively.

2. Related Work

Muhammad Omer, Thomas Kunz [4] have been investigated the most widely used operating systems for WSNs. Also understand the characteristics of popular OSs for WSN in particular and embedded devices in general without any studies about the relation between the OS and specific application.
Wei Dong, Xue Liu [5] have been examined the challenges of the OS design space. Then introduce constitutes a sensornet OS by describing its major components. Next, they provide an overview of the existing work, present a taxonomy of state-of-the-art sensornet OSes, and discuss various approaches to address the design challenges.
D. Manjunath [6] has been presented a well-rounded review of four popular operating systems proposed for WSNs: TinyOS, SOS, MANTIS, and Contiki. Inspired by the engineering approach, he has been identified the fundamental challenges involved in designing each component/feature of a typical sensorOS, and then described how these fundamental challenges have been approached by different sensor operating systems.
Hyunhak kim, Seongki [7] have been proposed an evolvable operating system for WSNs. Each component in this architecture is designed to perform its functionality concerned with power consumption, highly limited resources.
Margi, C.B, Escola de Artes [8] have been developed a comparison between two different operating systems which are Contiki and TinyOS that are running on the same hardware platform Crossbow TelosB. Using a set of tasks, which includes sensing, communication and security mechanisms, they have been evaluated their behavior in terms of energy consumption and execution time.
Ramon Serna, Ivan Shcherbakov, and others [9] have been presented an Operating System Abstraction Layer (OSAL) to reduce the portability efforts of software applications between platforms, independent of the running OS. They claimed that such design reduces dramatically the required efforts related to portability of application code, and effectively enables the re-utilization of software components in later deployments.
In this paper, mapping wireless sensor network applications requirements to existing operating systems will be introduced.

3. Design Issues and Challenges

WSN operates at two levels [10]. One is at the network level and the other is at node level. Network level interests are connectivity, routing, communication channel characteristics, and protocols. Node level interests are hardware, radio, CPU, sensors and limited energy. At a higher level OS for WSN can also be classified as node-level (local) and network-level (distributed). The important issues related to node-level are limited resource management, concurrency handling, power management and memory management where as issues related to both are inter-node communication, failure handling, heterogeneity and scalability.
This section discusses the important issues of both node and network-level to be considered while designing an operating system for WSN. These issues discuss the challenges and motivate the design requirements of an operating system needed for WSN.

3.1 Restricted Resource
A typical sensor node shown in figure 1 is constrained by the resources available to it. It is constrained by limited battery power, processing capability, memory and bandwidth.

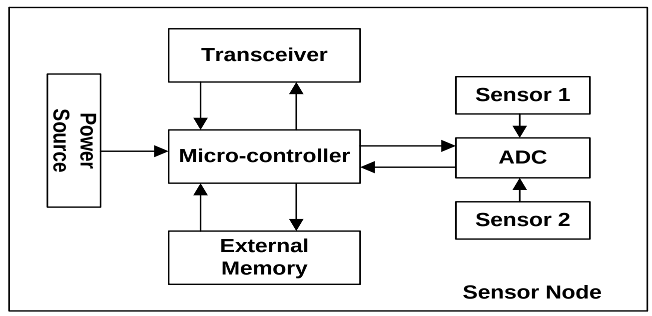


Fig. 1 Sensor Node Architecture

Figure 2 depicts, where operating system stands in the software layers of the WSN. Middleware and application layers are distributed across the nodes. Core kernel of the operating system sits at each individual node. On top of it, middleware and applications run as interacting modules across nodes.

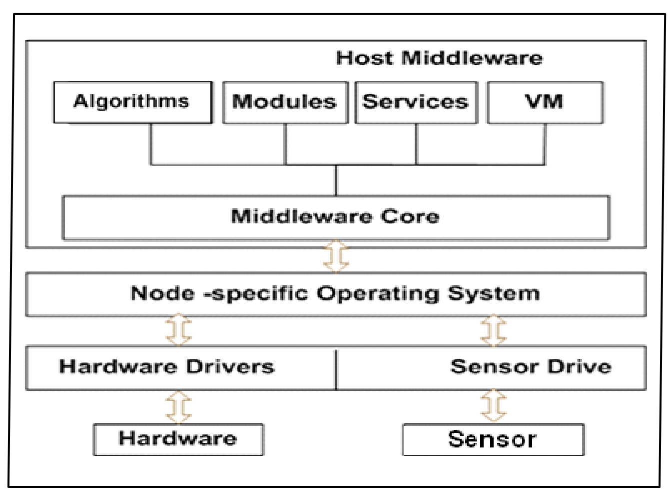


Fig. 2 Software Layers

ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

3.1.1 Battery power

Power consumption is crucial to the life span of WSN based applications. Most of the applications in WSN are long lived ranging from days to years. So a typical node with a limited power supply has to live mostly for months to years. Unlike conventional systems where the power is not at all a constraint factor in building the system, operating systems for sensor nodes have to consider power as one of the available resources like processor and memory.

3.1.2 Processing Power

Sensor nodes will have a processing power in the order of a few MIPS. Computation intensive operations should be properly scheduled; otherwise high priority tasks get delayed/starved. Computation models like event driven will follow run-to-completion model. This takes more processor time if the task is running for long time and preventing other jobs to wait for longer time irrespective of their priorities. Hence operating system should properly schedule the processor according to the priority of jobs.

3.1.3 Memory

The current generation of micro-controllers family such as Mica [11], its successors and some microcontrollers (e.g. nymph, EYES etc) specific to various research projects have nearly 128kbytes of program memory. One of the main constraints for the developer is this available program memory and operating system developed for WSN should fit within this memory. The system software such as operating system, virtual machine, middleware, and application algorithms have to fit into this memory. Optimal usage of this memory should start from lower level (i.e. Operating System).

3.1.4 Bandwidth

A typical sensor node uses RF channel to communicate with other sensor nodes in the network. ZigBee [12] is the emerging standard to define the communication protocol stack based on the existing physical and data-link layers of IEEE 802.15.4 Personal Area Network (PAN) standard. Data rate supported by PANs is 256kbps. Whereas Bluetooth standard supports data rate up to 3Mbps. CC1000 is another standard that has been widely used in sensor networks. Its data rate is around 39kbps.

3.2 Portability

The hardware platforms in WSN are evolving day-by-day. Portability is an important issue to be considered as everyone is working on their customized hardware platforms. Portability is one of the main concerns for the developer to make the software work on different hardware platforms. The operating system should be written in such a way that it is easily portable to different hardware platforms with minimal changes.

3.3 Customizability

Applications in WSN are spread over different disciplines. Specific applications of WSN include but not limited to monitoring environment, surveillance, target tracking etc. Survey of some of these applications can be found in [12] [13] [14]. Most of the software platforms developed for WSN are application specific. Different applications demand different requirements from operating system. These requirements may be reconfigurability, real-time guarantees. The design of OS should be in such a way that it should be easily customizable and extensible to various applications.

3.4 Multicasting

At a given point of time, nodes in the WSN could be doing more than one task. For example, consider a typical application where in the sensed data from the environment is collected, aggregated based on some filtering conditions, encrypted/decrypted and passed it towards the sink node through other nodes. In this application the sensor node has to do the following tasks at a given point of time:

- Sense the data.
- Collect data from other neighborhood sensor nodes.
- Aggregate the data based on the certain conditions provided.
- Route the data to the sink node.

3.5 Network Dynamics

Mobility, failure of communication channels/nodes constitutes the dynamics in WSN. Topologies are more prone to changes due to these dynamics which may result in network partitions. Link failures and the interferences in the RF communication channel deviates the behavior of the WSN from its normal operation. Operating system should adapt the application according to the context of different dynamics of the environment. This helps in providing transparency from network dynamics to the application.

3.6 Distributed Nature

There is a clear distinction between the services that should be supported by middleware and OS in traditional systems. This is masked in WSN due to cross layer interaction support which is a prominent feature for these kinds of systems. Sensor nodes in WSN are loosely coupled and sometimes deployed across a large geographical area. The scale of the network sometimes is in the order of thousands of sensor nodes. Each individual node has its own processing power, system software to run and the co-operation among the nodes happen through exchange of messages.

3.6.1 Heterogeneity

Heterogeneity in the network arrives due to varying level of node capabilities. This causes different nodes to be present in the network with different capabilities. These capabilities can be in terms of memory, sensing modality or residual energy, software components residing at the node. Many of the practical sensor networks are heterogeneous [14] in their sensing capability.

3.6.2 Scalability

Scalability here refers to the size of the network [15]. As the system is composed of large number of nodes, the system algorithms should work with acceptable performance degradation with increase in the number of nodes. In WSN there are subtle differences related to some issues in designing middleware and operating system. Some of the above issues seems to look like middleware issues but virtual machine approaches [12] and distributed operating systems [16] in WSN did concerned about them in designing.

4. Designed Characteristics

The following are the important design characteristics to be considered while designing an operating system for WSN.

4.1 Architecture

The organization of an OS constitutes its structure. The architecture of an OS has an influence on the size of the OS kernel as well as on the way it provides services to the application programs. Some of the well known OS architectures are the monolithic architecture, the micro-kernel architecture, the virtual machine architecture and the layered architecture.

A monolithic architecture [9] in fact does not have any structure. Services provided by an OS are implemented separately and each service provides an interface for other services. Such an architecture allows bundling of all the required service together into a single system image, thus results in a smaller OS memory footprint. An advantage of the monolithic architecture is that the module interaction costs are low.

An alternate choice is a microkernel architecture [9] in which minimum functionality is provided inside the kernel. Thus, the kernel size is significantly reduced. Most of the OS functionality is provided via user-level servers like a file server, a memory server, a time server, etc. If one server fails, the whole system does not crash. The microkernel architecture provides better reliability, ease of extension and customization.

A virtual machine [9] is another architectural choice. The main idea is to export virtual machines to user programs, which resemble hardware.

4.2 Efficient Execution Model

The execution model provides the abstraction of computational unit and defines services like synchronization, communication, and scheduling. These abstractions are used by the programmer for developing applications. Communication service defines the way the computational units communicate. They communicate to exchange data, delegation of functionalities and signaling. While communicating there can be data that is shared. Accessing shared data requires proper synchronization mechanisms to avoid race conditions.

4.3. Communication Protocol Support

In the OS context, communication refers to inter-process communication within the system as well as with other nodes in the network. WSNs operate in a distributed environment, where senor nodes communicate with other nodes in the network. All WSN OSs provide an Application Programming Interface (API) that enables application program to communicate. It is possible that a WSN is composed of heterogeneous sensor nodes, therefore the communication protocol provided by the OS must also consider heterogeneity.

4.4. Programming Model

The programming model supported by an OS has a significant impact on the application development. There are two popular programming models provided by typical WSN OSs, namely: event driven programming and multithreaded programming. Multithreading is the application development model most familiar to programmer, but in its true sense rather resource intensive, therefore not considered well suited for resource constraint devices such as sensor nodes. Event driven programming is considered more useful for computing devices equipped with scarce resource but not considered convenient for traditional application developers. Therefore researchers have focused their attention on developing a light-weight multithreading programming model for WSN Oss.

4.5. Scheduling

The Central Processing Unit (CPU) scheduling determines the order in which tasks are executed on a CPU. In traditional computer systems, the goal of a scheduler is to minimize latency, to maximize throughput and resource utilization, and to ensure fairness. The selection of an appropriate scheduling algorithm for WSNs typically depends on the nature of the application. For applications having real-time requirements, real-time scheduling algorithm must be used. For other applications, non-real-time scheduling algorithms are sufficient. WSNs are being used in both real-time [17] [18] and non-real-time [19] environments; therefore a WSN OS must provide scheduling algorithms that can accommodate the application requirements.

4.6 Resource Management
One of the fundamental tasks of an operating system is to manage the system resources efficiently. Resources available in a typical sensor node are processor, program memory, battery, and sensors etc. Efficient use of processor involves using a scheduler with optimal scheduling policy. Usage of memory involves memory protection, dynamic memory allocation, etc. Battery should be treated as a special resource. Sleep modes help in power management of battery. Managing sensors include controlling sensing rate.

5. Classification framework for WSN Operating Systems

Architecture, execution model, scheduling, routing protocols, hardware support, and application support have been chosen as the important design features that forms basis for our classification framework.
A different type of each feature is depicted in the figure 3. Based on each feature certain operating system, suitable hardware, and suitable routing protocol could be suggested. Miscellaneous features shown in the figure are explained for each operating system. These features are simulation support and programming language. Below is an overview of the design features.

5.1. Classification According to Architecture
Architecture of the kernel influences the way it provides services as described in section 4.1.
Operating systems which fall under monolithic is TinyOS [16]. Also suitable hardware for monolithic is Mica2. Suitable routing protocol for monolithic is Sensor Protocol for Information Negotiation (SPIN). While operating systems which fall under modular is LiteOS [20].
Also suitable hardware for modular is Micaz. Suitable routing protocol for modular is MAC protocol. Operating systems which fall under VM is ContikiVM [21]. Also suitable hardware for VM is MSP430. Suitable routing protocol for it is multi-hop mesh routing as shown in figure 3. Architecture for different OS described in figure 4.

5.2 Classification According to Execution Model
Operating systems which fall under event-based is SOS [10]. Also suitable hardware for it is Tmote-sky. Suitable routing protocol for event-based execution model is directed diffusion. While operating systems which fall under thread-based is MantisOS [22]. Also suitable hardware for it is Mica2. Suitable routing protocol for thread-based is Sequential Assignment Routing (SAR). Operating systems which fall under Hybrid is Contiki [21]. Also suitable hardware for Hybrid is Avr MCU. Suitable routing protocol for it is Gradient-Based Routing (GBR) as shown in figure 3. Execution model of different Oss described in figure5.

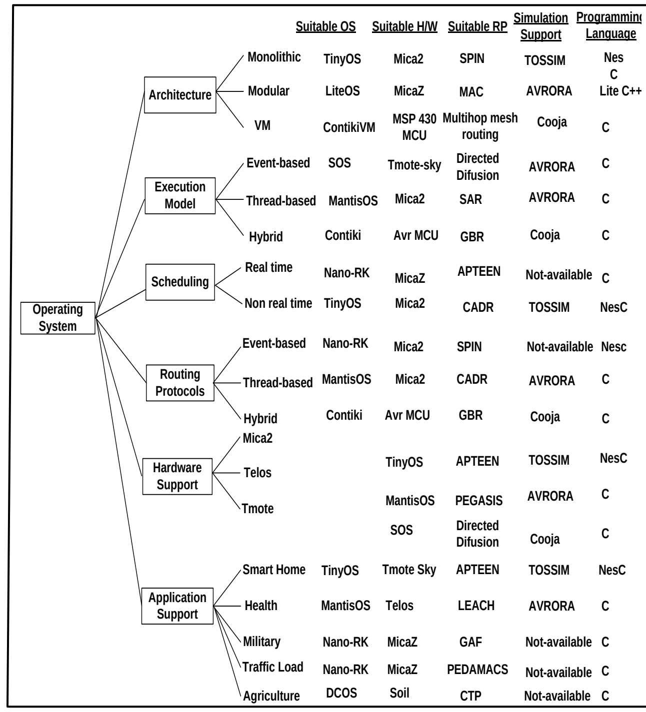


Fig. 3 Classification Framework for WSN operating system

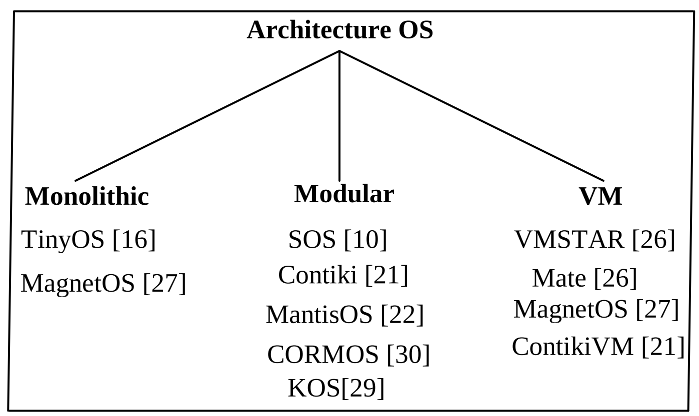


Fig. 4 Architecture of Different OSs

5.3. Classification According to Scheduling
Operating systems which fall under real time scheduling is Nano-RK [17]. Also suitable hardware for it is MicaZ.
Suitable routing protocol for real time scheduling is Threshold Sensitive Energy Efficient Sensor Network Protocol (APTEEN). While operating systems which fall under non-real time scheduling is TinyOS[16]. Also suitable hardware for it is Mica2. Suitable routing protocol for non-real time scheduling is Constrained Anisotropic

diffusion routing (CADR) as shown in figure 3. Read time and non-real time OS classification described in figure 6.

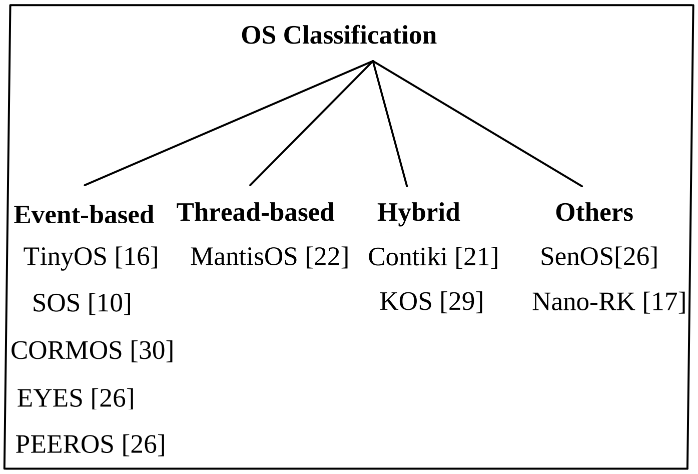


Fig. 5 Execution Model of Different OSs

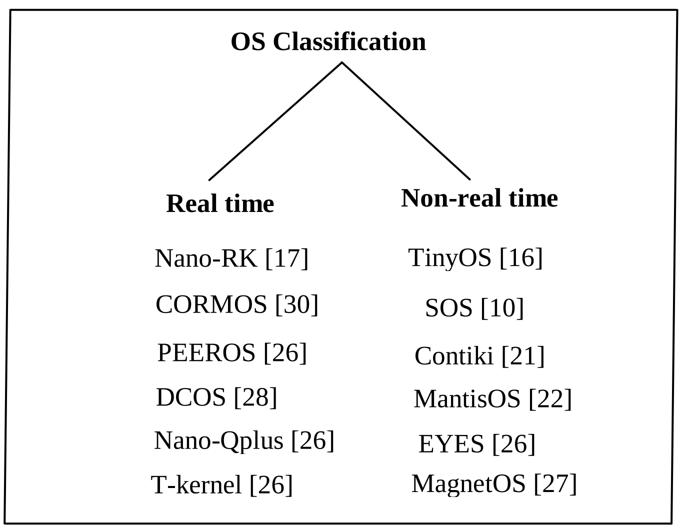


Fig. 7 Real-time and Non-real-time OS Classification

5.4. Classification According to Routing Protocols
Routing protocols can be classified into event based routing protocols [37], thread based routing protocols [38], and hybrid routing protocols [39]. The Example for event based is directed diffusion where the network will be active only when an event occur, otherwise the network is idle. The example for thread based is CADR. Example for hybrid routing protocol is GBR. Figure 3 presents the suitable hardware and operating system for each type of routing protocol.
Operating systems which fall under event-based is Nano-RK [17]. Also suitable hardware for it is Mica2. Suitable routing protocol for event-based is Sensor Protocol for Information Negotiation (SPIN). While operating systems which fall under thread-based is MantisOS [22]. Also suitable hardware for it is Mica2. Suitable routing protocol for thread-based is CADR. Operating systems which fall under Hybrid is Contiki [21]. Also suitable hardware for Hybrid is Avr MCU. Suitable routing protocol for it is GBR as shown in figure 3.

5.5. Classification According to Hardware Support.
There are many types of hardware WSN. Each type of hardware has suitable operating system. So the first thing to design any application is choose suitable sensor for application target. For example in real application the type of hardware differs from non-real applications.
Operating systems which fall under Mica2 hardware is TinyOS [10]. Also suitable routing protocol for it is APTEEN. While operating systems which fall under Telos hardware is MantisOS [22]. Suitable routing protocol for it is Power efficient Gathering Sensor Information System (PEGASIS).Operating systems which fall under Tmote hardware is SOS [10]. Also suitable routing protocol for it is directed diffusion as shown in figure 3. The lists of platforms supported as of now are shown in figure 7.

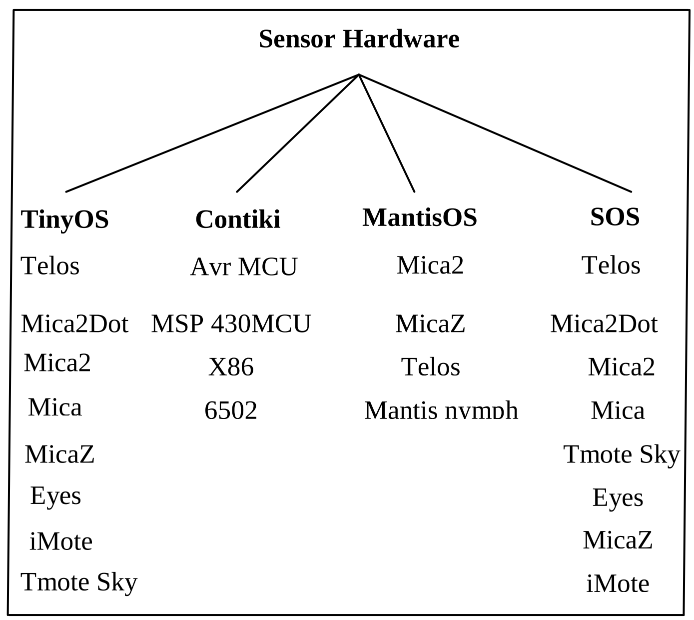


Fig. 7 Hardware Platforms Supported by different OS

5.6.Classification According to Application Support
Classification of applications depends on the nature of the purpose they are used for. In environment monitoring, the application can be broadly classified into indoor and outdoor monitoring [14]. While in Health application Involves monitoring patients and alerting doctors. Here sensors measure the recent actions of the patients and the Remind the doctors about the behavior of the patient.

Category	Suitable H/W	Suitable OS	Suitable RP	Topology	Programming Language
Environment Monitoring	Tmote	TinyOS	DD	Multi-hop	nesC
Health	Telos	MantisOS	LEACH	Cluster-head	C
Military	MicaZ	Nano-RK DCOS	GAF	Multi-hop	C
Agriculture	Soil sensor	Nano-RK	CTP	Tree-topology	C
Smart Home	Tmote Sky	TinyOS	APTEEN	Three tier	nesC
Traffic Load	MicaZ	T-kernel	PEDAMACS	Multi-hop	C
Habitat Monitoring	Mica2	SOS	SPAN	Cluster-head	C
Production/ Commercial	Telos	TinyOS	SAR	Three tier	nesC

Table 1: Application Characteristics and Suitable OSs

Another application might be tracking patients, doctors, and drug usage in the hospitals. In military application, Sensor nodes are well suited to the need of military applications. Interesting of them are information collection, enemy tracking, battlefield surveillance, target classification, perimeter security, border patrol. Other applications are smart home [23], Agriculture [24], and Traffic Load [32].
Characteristics of applications are evaluated against the categories of applications, suitable hardware, suitable operating system, suitable routing protocol, type of topology and programming language as shown in Table 1.
Suitable hardware which falls under environment application is Tmote. Also suitable OS for it is TinyOS. Suitable routing protocol for it is Directed Diffusion (DD). Type of topology for this application is multi-hop. Suitable hardware which falls under health application is Telos. Also suitable OS for it is MantisOS. Suitable routing protocol for it is LEACH. Type of topology for this application is cluster-head. Suitable hardware which falls under military application is MicaZ. Also suitable OS for it is DCOS [26]. Suitable routing protocol for it is Geographic Adaptive Fidelity (GAF). Type of topology for this application is multi-hop as shown in Table 1.
Suitable hardware which falls under agriculture application is Soil sensor. Also suitable OS for it is Nano-RK. Suitable routing protocol for it is Collection Tree Protocol (CTP). Type of topology for this application is tree-topology.
Suitable hardware which falls under smart home application is Tmote. Also suitable OS for it is TinyOS. Suitable routing protocol for it is Adaptive Threshold Sensitive Energy Efficient Sensor Network Protocol (APTEEN). Type of topology for this application is three-tier.
Suitable hardware which falls under traffic load application is MicaZ. Also suitable OS for it is T-kernel [26]. Suitable routing protocol for it is Power Efficient and Delay Aware Medium Access Protocol (PEDAMACS). Type of topology for this application is multi-hop.
Suitable hardware which falls under habitat application is Mica2. Also suitable OS for it is SOS. Suitable routing protocol for it is SPAN. Type of topology for this application is cluster-head.
Suitable hardware which falls under production application is Telos. Also suitable OS for it is TinyOS. Suitable routing protocol for it is SAR. Type of topology for this application is three-tier as shown in Table 1.

6. Comparative analysis of Operating Systems

Summary of the existing operating systems such as TinyOS, SOS, Contiki, Mantis, and others presented in Table 2. The supported Features by them such as execution model, real time guarantee, dynamic programming, priority-based scheduling, and application supported are shown in Table 2.
Execution model which fall under TinyOS is component-based. Which no have real time guarantee, priority scheduling, and dynamic programming. Suitable application for it is smart home.
In another type, Execution model which fall under SOS is module-based. Which have priority scheduling, and dynamic programming. Suitable application for it is habitat monitoring. Also, Execution model which fall under Contiki is Hybrid. Which no have real time guarantee, and priority-based scheduling. Suitable application for it is temperature and humidity.

OS	Execution Model	Real Time Guarantee	Dynamic Reprogramming	Priority-based Scheduling	Application
TinyOS [16]	Component Based	No	No	No	Smart Home [23]
SOS[10]	Module based	No	Yes	Yes	Habitat monitoring [33]
Contiki [21]	Hybrid	No	Yes	No	Temperature and humidity [35]
Mantis [22]	Thread based	No	Yes	No	Health [31]
EYES [26]	Event Based	No	No	Yes	Light [25]
Nano-RK [17]	Task Based	Yes	Yes	Yes	Agriculture [24]
DCOS [28]	Data centric	Yes	Yes	No	Military [34]
MagnetOS [27]	VM based	Yes	No	No	Biomedical [36]
SenOS [26]	State based	No	Yes	No	Time-critical application [34]
T-kernel	Task Based	Yes	Yes	Yes	Traffic Control [32]

Table 2: Summary of Operating Systems

In MantisOS, Execution model which fall under it is thread-based. Which no have real time guarantee, and dynamic programming.
Suitable application for it is health. In EYES, Execution model which fall under it is event-based. Which no have real time guarantee, and priority-based scheduling. Suitable application for it is light. Execution model which fall under Nano-RK is task-based. Which have real time guarantee, and dynamic programming. Suitable application for it is agriculture.
In another type, Execution model which fall under DCOS is data-centric. Which have priority scheduling, and dynamic programming. Suitable application for it is military. In MagnetOS, Execution model which fall under it is VM-based. Which no have dynamic programming, and priority-based. Suitable application for it is biomedical. Finally, Execution model which fall under SenOS it is state-based. Which no have real time guarantee, and priority-state. Suitable application for it is time-critical.

7. Conclusions

This paper studies in depth the different operating systems related to WSN, then it maps between the WSN applications and the suitable OS according to the application's requirements. The contribution of this work is two-fold, the first contribution of this work is deriving and forming a classification framework for the existing OSs According to a lot of noteworthy important OS features, such as, architecture, Execution Model, Scheduling, and Routing Protocols, this classification helped in understanding the contrasting differences among existing operating systems, and lays a foundation to design an ideal WSN OS. The second contribution is offering a guide for selecting the appropriate OS to the desired WSN application where also the existing WSN applications have been classified and the applications categories have been mapped to the OS features. For example operating system which falls under agriculture application is Nano-RK, Whereas, the suitable OS for the time critical application is SOS. This helps the application developer in choosing the appropriate OS, based on the application requirements.

8. Future Work

Because of the potential applications of WSN, and growing interest on WSNs enforced to deploy multiple applications simultaneously on the same network. It gives an advantage of reducing infrastructure deployment cost. But, it is an interesting research challenge. Till now, there are not any OSs that facilitate running multiple applications. Though other desirable characteristics like low memory foot print and portability are mandatory, the above are the important objectives to be considered while designing new OS for an embedded WSN systems.

References

[1] Zhang, Junqi, Varadharajan, and others, “Wireless sensor network key management survey and taxonomy”, Journal of network and computer applications, Vol. 33, Issue 2, pp.63-75, 2010.

[2] Basma M. Mohammad El-Basioni, Sherine M. Abd El-kader, Hussein S. Eissa, and Mohammed M. Zahra, “Clustering in Wireless Sensor Network: A Study on Three Well-known Clustering Protocols” Accepted as a book chapter in Developments in Wireless Network Prototyping, Design and Deployment: Future Generations, M. A. Matin (Ed.), IGI Global, 2013.

[3] Basma M. Mohammad, Sherine M. Abd El-kader, and Hussein S. Eissa, “Designing a local path repair algorithm for directed diffusion protocol”, Egyptian Informatics Journal, Vol. 13, No. 3, pp. 155–169, November 2012.

[4] Muhammad Omer Farooq, and Thomas Kunz, “Operating Systems for Wireless Sensor Networks: A Survey “, MDPI Open Access Information and Policy, Vol. 11, No. 6, 2011.

[5] Wei Dong, and Xue Liu, “Providing OS Support for Wireless Sensor Networks: Challenges and Approaches”, IEEE Communication Surveys and Tutorials, Vol. 12, No. 4, November 2010.

[6] D. Manjunath, “A Review of Current Operating Systems for Wireless Sensor Networks”, In Proceedings of the 3th European Conference on Wireless Sensor Networks, Vol. 2, No. 1, pp. 293-308, January 2006

[7] Thu-Thuy, Do, Daeyoung Kim, Tomás Sánchez López, and Hyunhak Kim, “An Evolvable Operating System for Wireless Sensor Networks“, International Journal of Software Engineering and Knowledge Engineering, Vol. 15, No. 1 , 2009.

[8] Margi, Escola de Artes, Cienc. E Humanidades, and others, “Impact of Operating Systems on Wireless Sensor Networks (Security) Applications and Testbeds”, In Proceedings of the 19th International Conference on Computer Communications and Networks, Vol. 6, No. 2, , 2010.

[9] Ramon Serna, van Shcherbakov, and Gerhard Fohler, “An Operating System Abstraction Layer for Portable Applications in Wireless Sensor Networks”, SAC ‘10 Proceedings of the 2010 ACM Symposium on Applied Computing, Vol. 10, No. 1, 2010.

[10] Shripad V Deshpande, and P. R. Devale, “Recent Trends in Using Wireless Sensor Networks in Industrial Environment”, International Journal of Computer Networking, Vol. 3, No. 3, 2013.

[11] Hyoseung, and Hojung Cha, “Multithreading Optimization Techniques for Sensor Network Operating Systems”. In Proceedings of the 4th European Conference on Wireless Sensor Networks, Vol. 4, No. 2, pp. 293-308, January 2007.

[12] M. Hussnain, M. Sharjeel, S. R. Chaudhry, S. A. Hussain, and others, “Investigating Multi-Topological ZigBee Based Wireless Sensor Network in Precision Agriculture”, Journal of Basic and Applied Scientific Research, Lahore, Pakistan, Vol.3, No. 2, 2013.

[13] Basil Hamed, “Design and Implementation of Smart House Control Using Lab VIEW”, International Journal of Soft Computing and Engineering (IJSCE), Vol. 1, No. 6, January 2012.

[14] Mohammed A. Hussein, “A Novel Approach for Indoor Outdoor Air Pollution Monitoring”; IJCSI International Journal of Computer Science Issues, Vol. 9, No. 4, University of Sulaimani, July 2012.

[15] Chris Karlof, Naveen Sastry , and David, “TinySec: A Link Layer Security for Wireless Sensor Networks”, In Proceedings of the 2th ACM SenSys, Baltimore, MD, Vol.3, No. 5, November 2004.

[16] Cooprider, Archer, Eide, and others, “Efficient Memory Safety for Tinyos”. In Proceedings of the 5th International Conference on Embedded Networked Sensor Systems (SenSys’07), Vol.4, No. 3, pp. 205-218, New York, NY, USA, November 2007.

[17] Eswaran, Rowe, Rajkumar, and others, “Nano-RK: An Energy-Aware Resource-Centric RTOS for Sensor Networks”. In Proceedings of the 26th IEEE Real-Time Systems Symposium, Miami, USA, Vol. 2, No. 1, December 2005.

[18] Rowe, Mangharam, Rajkumar, and R. FireFly, “A Time Synchronized Real-Time Sensor Networking Platform”. In Proceedings of the 28th IEEE Real-Time Systems Symposium, Tucson, USA, Vol. 3, No. 7, 2007.

[19] Rowe, Lakshmanan, and Yhu, Rajkumar, “Rate-Harmonized Scheduling for Saving Energy”. In Proceedings of the 29th IEEE Real-Time Systems Symposium, Barcelona, Spain, Vol.2, No. 2, 2008.

[20] Cao, Abdelzaher, Stankovic, and others “The LiteOS Operating System: Towards UNIX like Abstraction for Wireless Sensor Networks”. In Proceedings of the 7th International Conference on Information Processing in Sensor Networks (IPSN), USA, Vol. 5, No. 2, April 2008.

[21] Adam Dunkels, Bjorn Gronvall, and Thiemo Voigt, “Contiki – A Lightweight and Flexible Operating System for Tiny Networked Sensors”, In Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks, Vol.2 No. 1, 2004.

[22] Bhatti, Carlson, Dai, Deng, and others, “Mantis OS: An Embedded Multithreaded Operating System for Wireless Micro Sensor Platforms”, Mobile Network. Appicationl, Vol. 10, Issue 2, pp. 563-579, 2005.

[23] Basma M. Mohammad , Sherine M. Abd El-kader, and Mahmoud Abdelmonim Fakhreldin, “Smart Home Design using Wireless Sensor Network and Biometric Technologies”, International Journal of Application or Innovation in Engineering and Management (IJAIEM), Volume 2, Issue 3, pp. 413–429, March 2013.

[24] Sherine M. Abd El-kader and Basma M. Mohammad El-Basioni, “Precision Farming Solution in Egypt Using the Wireless Sensor Network Technology”, Egyptian Informatics Journal, Volume 14, Issue 3, July 2013.

[25] A. Mason1, L. E. Cordova Lopez, and A. Shaw, “Road traffic management using GIS and WSN”, Built Environment and Sustainable Technologies, Liverpool John Moores University, Liverpool, UK, Vol.4, Issue 2, 2011.

[26] A. K. Dwivedi, M. K. Tiwari, and O. P. Vyas, “Operating Systems for Tiny Networked Sensors: A Survey”, International Journal of Recent Trends in Engineering, Vol. 1, Issue 2, May 2009.

[27] J.Radhika, and S.Malarvizhi, “Middleware approaches for Wireless Sensor Networks: An overview”, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 6, No 3, November 2012.

[28] T. Hofmeijer, S. Dulman, P. G. Jansen, and P. J. M. Havinga. “Dcos, a real-time light-weight data centric operating system”. Int. Conf. on (ACST), St. Thomas, Virgin Islands, USA, pp. 259-264, Calgary, Canada, November 2004.

[29] Yan Liang; Jiannong Cao; Zhang, D.; Rui Wang; Quan Pan “A Biologically Inspired Sensor Wakeup Control Method for Wireless Sensor Networks”, Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, On pp. 525 – 538 Volume: 40, Issue: 5, Sept. 2010.

[30] J. Yannakopoulos and A. Bilas. Cormos: A communication-oriented runtime system for sensor networks. In The Second European Workshop on Wireless Sensor Networks (EWSN 2005), February 2005.

[31] J. Stankovic, Q. Cao, T. Doan, L. Fang, Z. He, R. Kiran, S. Lin, S. Son, R. Stoleru, and A. Wood. “Wireless sensor networks for in-home healthcare: Potential and challenges”. In High Confidence Medical Device Software and Systems (HCMDSS) Workshop, Philadelphia, PA, June 2005.

[32] Maha Mohamed Nabeel, Mahmoud Fakher el Deen, Sherine Abd El-Kader, “Intelligent Vehicle Recognition based on Wireless Sensor Network,” International Journal of Computer Science Issues (IJCSI), Volume 10, Issue 4, July 2013

[33] Tomasz Naumowicz, Robin Freeman, and others, “Wireless Sensor Network for habitat monitoring on Skomer Island”, In proceeding of: The 35th Annual IEEE Conference on Local Computer Networks, LCN 2010, 10-14 October 2010, Denver, Colorado, USA.

[34] Tian He, John A. Stankovic, and others, “Energy-efficient surveillance system using wireless sensor networks”, In Proceedings of the 2nd international conference on Mobile systems, applications, and services, pp.270-283, ACM New York, NY, USA, 2004.

[35] Kshitij Shinghal 1, Arti Noor, and others, “Intelligent Humidity Sensor for Wireless Sensor Network Agriculture Application”, International Journal of Wireless and Mobile Networks (IJWMN) Vol. 3, No. 1, February 2011.

[36] Loren Schwiebert, Sandeep K.S, and others, “Research challenges in wireless networks of biomedical sensors”, In Proceedings of the 7th annual international conference on Mobile computing and networking, pp. 151-165, ACM New York, NY, USA, 2001.

[37] Anar A. Hady *, Sherine M. Abd El-kader, Hussein S. Eissa, “Intelligent Sleeping Mechanism for wireless sensor networks”, Egyptian Informatics Journal, ISSN 1110-86652013, April 2013.

[38] T.salem, Salah.Abd-Elmagied, and others “Improving the Routing Mechanism of Clusters Sensor Networks”, International Journal of Intelligent Computing and Information Science, Vol.10, Issue 3, July 2012.

[39] Venugopalan Ramasubramanian , Zygmunt J. Haa, “SHARP: A Hybrid Adaptive Routing Protocol for Mobile Ad Hoc Networks”, In Proceedings of the 4th ACM international symposium on Mobile ad hoc networking & computing, pp. 303-314, ACM New York, USA, 2003.

AUTHOR

Tarek M. Salem is an assistant researcher in Computers and Systems Department at the Electronics Research Institute (ERI) in Egypt. In May 2013, he completed her B.S. in Computers and Systems department, Faculty of Engineering, Al-Azhar University. During the 2005-2012 year, he joined the Arabic Organization for Industrialization. In 2013 year, he occupied the position of research assistant at Electronics Research Institute. He is a Lecturer, at Aljazeera Institute for information technology from 2011 till now. Now, he is studying for Ph.D. degree.

S. Abd El-kader has her MSc, & PhD degrees from the Electronics & Communications Dept. & Computers Dept., Faculty of Engineering, Cairo University, at 1998, & 2003. Dr. Abd El-kader is an Associate Prof., Computers & Systems Dept., at the Electronics Research Institute (ERI). She is currently supervising 3 PhD students, and 10 MSc students. Dr. Abd El-kader has published more than 25 papers, 4 book chapters in computer networking area. She is working in many computer networking hot topics such as; Wi-MAX, Wi-Fi, IP Mobility, QoS, Wireless sensors Networks, Ad-Hoc Networking, realtime traffics, Bluetooth, and IPv6. She is an Associate Prof., at Faculty of Engineering, Akhbar El Yom Academy from 2007 till 2009. Also she is a technical reviewer for many international Journals. She is heading the Internet and Networking unit at ERI from 2003 till now. She is also heading the Information and Decision making support Center at ERI from 2009 till now. She is supervising many automation and web projects for ERI. She is supervising many Graduate Projects from 2006 till now. She is also a technical member at both the ERI projects committee and at the telecommunication networks committee.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 1, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

Challenges and Opportunities for the Implementation of Social Network Technologies (SNTs) in Teaching in Universities in Ghana

Kofi Sarpong Adu-Manu[1], John Kingsley Arthur [2] and Clement Yeboah[3]

[1] Computer Science Department, Valley View University
Accra, Ghana

[2] Computer Science Department, Valley View University
Accra, Ghana

[3] Department of Business Administration , Strayer University, Columbus Campus
Columbus, Ohio, USA

Abstract

The introduction of Information and Communication Technology (ICT) has changed the mode of delivery in many higher institutions and enhance student participation. In most developing countries however, ICT has not pervaded. This study is descriptive, which adopts a case study approach to investigate the challenges and opportunities in the implementation of Social Network Technologies (SNTs) in four private universities in the Greater Accra Region, Ghana. Eighty (80) lecturers in private universities were used as the sample for the study. A well-developed tool on the implementation challenges and opportunities of SNTs was used for data collection. The instrument contained twenty three (23) listings/items. The data collected were analyzed using frequency distribution and mean. From the research the findings reached with regard to the challenges which are higher cost of licensing the SNTs and purchase of sophisticated devices, incompetence in the use of multimedia tools, safety and privacy issues, and change in teaching methodology brings about dissatisfaction and the notion of uncertainty due to lack of exposure to technological devices. The opportunities available with SNTs for teaching are that SNTs support different teaching methods due to their flexibility, creation of effective teaching and learning environment and making teachers location independent. The following recommendations are made management needs to train its faculty on how to use cutting edge technologies such as SNTs to be implemented in teaching. University management should consider how to make laptops available for all of their lecturers so that they will be encouraged to have a strong affiliation with the technology and they will overcome the lack of exposure and change their teaching methods and move away from the traditional methods of the teaching. It will also boost their frequency of usage.

Keywords: *Social Network Technologies, information technology, ICT Infrastructure, multimedia tools.*

1. Introduction

Advances in technology have over the years seen a drastic improvement in the development of highly sophisticated devices (mobile) and other tools which can be adopted in enhancing teaching. According to Jung (2008) technology has the potential of improving quality of education, increasing access to education, pedagogical innovation and creating high market value.

In increasing access, ICT expands the learning opportunities as people will be able to learn any time and everywhere. This will lessen the gap in education by reaching to remote, underdeveloped and marginalized population (Zamza, n.d.). Apart from the hardware infrastructure, there has been improvement in software development which seems to bring the globe closer and closer. Social networking facilities as an ICT tool have the potential to increasing performance of students when effectively used by lecturers.

Social networking has become one of the most important communication tools among people nowadays (Zamza, n.d.). Teachers are gaining interest in online education and students' participation is high. There are several challenges as well as opportunities with the use of technologies (e.g. social media tools) in teaching.
Generally in developing countries, access to internet is expensive, teachers lack technical and pedagogical skills to use technology and inadequate funds for staff and infrastructure development (Zamzam, n.d., & Chapman et al.2004). Amidst these challenges, Ghana has seen high

increase in the usage of mobile devices and other technological devices which both faculty members within the universities have access. According to Sife et al.(2007), the pedagogical and socio-economic forces that have driven the higher learning institutions to adopt and incorporate ICTs in teaching and learning include greater information access; greater communication; synchronous and asynchronous learning; increased cooperation and collaboration, cost-effectiveness and pedagogical improvement.

According to Turban et al. (2010), in their paper E-commerce a managerial Perspective stated that, "social networking is built on the idea of how people know and interact with each other". It gives people the power to share, making the world more open and connected. Nowadays, social networking has a vital influence on our live as it helps a lot in every field of life such as political field, economic field and educational field. However, this paper tries to highlight on the use of social networking in education and explain the advantages and disadvantages of using social networking for educational purpose (Zaidieh, 2012).

The use of social technologies for educational purposes has changed the demands and direction of higher education. Lecturers are now being encouraged to use social technologies in their teaching in order to encourage social learning and to prepare students as graduates who will contribute to a society that now relies heavily on social technologies (Hamid, 2011). It is against this background that the researchers intend to find out the challenges and opportunities available for the implementation of the social media technologies in the private universities in Ghana.

2 Statement of Problem

In Ghana and other developing countries, SNTs are challenged with the problem of internet facilities and other devices that contribute to the service. Emails have been used in China to facilitate communication among students themselves and their instructors, to submit assignments to their teachers and to post e-learning materials for students to read (Lee, 2004; & Guo & Cai, 2006).

The problem is that SNTs usage by lecturers in the universities (private or public) is challenged by the introduction of new technologies in terms of its availability and use amidst the challenges and the numerous opportunities these technologies offer to both the teacher and the learner. It is against this background that the study is carried out to determine the opportunities available to lecturers when using SNTs for teaching and secondly, the research seeks to identify the challenges faced by lecturers in adopting SNTs in teaching.

3. Objectives

This research examines

1. The challenges lecturers can face in implementing SNTs in teaching at the private universities.
2. The opportunities available for the use of SNTs in teaching in private/public universities.

4. Research questions

1. What challenges can lecturers face in the implementation of SNTs in teaching in the universities?
2. What opportunities are available to lecturers in implementing SNTs to improve teaching in the universities?

5. Review of Literature

Teachers often do not feel confident enough with their ICT skills to experiment with Learning 2.0 strategies (Redecker, 2010), which is one of the examples of social networking technologies. According to Conlon & Simpson as cited by Picardo (2011), early adopters see SNTs as tools that can help refine and develop teaching and learning. On the contrary, it is also perceived by teachers who remain skeptical about the adoption of emerging technologies because of the demands placed upon them for learning and understanding the new pedagogy involved and they often feel constrained by the contexts and pressures in which they work.

This will implicate the acceptance of the technology by teachers. According to Redecker (2010), embedding social media tools in education demands a change in the role of teachers, who have to act as guides and mentors, enabling and facilitating self-regulated learning processes. There are many uncertainties regarding the benefits of technology and the changes that the adoption of technology necessitates, such as the demand for technical support, pedagogical and instructional management issues, teacher professional development, network infrastructure, and costs of all components (Dooley, 1999).

According to Redecker (2010), Social media support more engaging and playful approaches, provide new formats for creative expression, and encourage learners and teachers to experiment with different, innovative, ways of articulating their thoughts and ideas. Using these social networking technologies such as twitter, Google+, WhatsApp, both the students and lecturers have an opportunity to create their own content. Posting on wall,

replying, commenting, and blogging aids creativity and innovations in both the students and the lecturers.
According to Zaidieh (2012), Social networking technologies can provide stronger understanding and increase retention on the subject, due to using many elements which exist under e-learning, e.g. multimedia, quizzes, interaction ... etcetera and the ability to retry training and over in order to understand. According to Zaidieh (2012) the networking potential of social media, together with its power to overcome time and space barriers, supports interaction and collaboration among and between learners and teachers who are geographically dispersed and enables students to broaden their horizons, and collaborate across borders, language barriers, and institutional walls.

6 Methodology

The research employed a survey research design. It is a descriptive research since it will purposely be used to describe, observe and document aspects of situation as it naturally occurs. It serves as a starting point for hypothesis generation or theory development. The population of the study was composed of all lecturers in four (4) selected universities in Ghana.

The sample size for the study was composed of eighty (80) lecturers who were selected from the total population of 300 lecturers. Two (2) sampling methods was deployed – the stratified sampling method and non-random sampling for effective selection. The research instrument used for the data collection was questionnaire designed for the respondents (lecturers). It was structured on a seven-point scale and has three sections (1-3) which sought information on the two (2) research questions.

The content validation of the instrument was established by three experts each in the area of Education, Information Technology and Business Administration respectively. The primary data for the study was obtained from respondents drawn from the population and were analyzed using frequency distribution and mean. Since the questions were rated on a seven-point scale, the researchers based on a mid-point scale of 3.50 as the ceiling value to make decisions. Mean values that fell below 3.50 were regarded as not agreeing to the factors itemized and mean values above 3.50 were considered as agreeing to the listings. The Statistical Package for Social Science (SPSS) was used to analyze the data.

7 Results/Discussions

The discussions here are based on the research questions. The first question states that; what challenges can lecturers face in the implementation of SNTs in teaching in the private universities? Whereas the second question looks at what opportunities are available to lecturers/teachers in implementing SNTs to improving teaching in the private universities? From the survey conducted, the results were tabulated and summarized in the tables 1 and 2.

The columns in Table 1 are labelled as follows: S/N represents the Item Number; Opportunities available for implementing SNTs describe the items (as identified by S/N) as possible challenges one can face in implementing SNTs; CA represents Completely Agree that it is a challenge; MA represents Moderately Agree that it is a challenge; SA represents Somewhat Agree that it is a challenge; NT represents Neutral that is not sure whether or not is a challenge; SD represents Somewhat Disagree that it is a challenge; MD represents Moderately Disagree that it is a challenge; CD represents completely disagree that it is a challenge, X represents the mean score of responses for each item. From the Table above, it shows that all the items (1 to 9) with mean ratings of 4.0375, 5.3125, 5.35, 3.8375, 4.6875, 4.475, 5.2125, 4.4375 and 4.0125 are challenges.

The columns in Table 2 are labelled as follows: S/N represents the Item Number; Opportunities available for implementing SNTs describe the items (as identified by S/N) as possible challenges one can face in implementing SNTs; CA represents Completely Agree that it is an opportunity; MA represents Moderately Agree that it is a challenge; SA represent Somewhat Agree that it is a challenge; NT represents Neutral that is not sure whether or not an opportunity; SD represents Somewhat Disagree that it is an opportunity; MD represents Moderately Disagree that it is an opportunity; CD represents completely disagree that it is an opportunity, X represents the mean score of responses for each item. From the Table above, it shows that all the items (10 to 14) with mean ratings of 4.7375, 5.1625, 5.00, 4.6875, and 5.45 are opportunities.

Table 1: Response to challenges in implementing SNTs in teaching

S/N	Possible Challenges in implementing SNTs	CA	MA	SA	NT	SD	MD	CD	X
1	Inadequacy/Lack of ICT resources/infrastructure will negatively affect the use of SNTs for teaching	8	17	2	22	18	0	13	4.0375
2	Incompetence in the use digital and multimedia tools will negatively affect the implementation of SNTs	24	8	23	19	6	0	0	5.3125
3	Costs for licensing SNTs for the purpose of utilizing their full features will negatively affect their implementation	16	21	30	9	0	0	4	5.35
4	Physically challenged lecturers may have difficulty in the usage of SNTs therefore negatively will affect the implementation of SNTs	8	16	0	33	0	0	23	3.8375
5	The use of SNTs will call for change in learning methodologies and will negatively affect their implementation	16	8	30	7	9	0	10	4.6875
6	The notion of uncertainty to the use of SNTs in teaching	16	8	30	7	9	0	10	4.475
7	Safety and privacy concern on the use of SNTs will negatively affect its implementation	27	3	10	40	0	0	0	5.2125
8	The lack of gestures as experienced in the use of SNTs may result in Miscommunication, thereby affecting their implementation	18	2	9	38	0	7	6	4.4375
9	The implementation of SNTs for learning may result in misappropriation of time and health hazards	6	9	18	27	0	7	13	4.0125

Table 2: response to opportunities available to lecturers

S/N	Opportunities available for implementing SNTs	CA	MA	SA	NT	SD	MD	CD	X
10	The use of SNTs for teaching through blogging, writing on wall, commenting, and replying, posting events will bring up creativity and innovation in teaching.	16	12	9	31	7	0	5	4.7375
11	SNTs can provide effective learning environment for studies, thereby resulting in quality academic outcomes because they have elements such as multimedia quizzes.	16	27	8	20	5	0	4	5.1625
12	SNTs will make teaching location independent	11	21	22	14	7	5	0	5.00
13	SNTs are convenient and easy to use	16	10	18	13	19	0	4	4.6875
14	SNTs are flexible and have support for different styles of teaching.	35	10	9	14	6	6	0	5.45

7.1 Findings on Research Question 1

The findings from research question 1 indicated in table 1 show that lecturers/teachers go through a great deal of challenge when it comes to implementing social network in teaching. There are many challenges to bringing social media to the learning environment. Some of them are technical and some of them are conceptual and cultural. In this implementation of a social learning environment several challenges have been discovered.

The challenge for educators is to adapt to these technologies correctly: taking advantage of their strengths and avoiding their weaknesses (Dooley, 1999). It is evident from the survey that there are several challenges lecturers/teachers face in the implementation of social network platforms such as:

1. Safety and privacy issues
2. Misuse of time
3. Lack of gesture resulting in miscommunication
4. The notion of uncertainty due to lack of exposure to SNTs
5. Incompetence in the use of multimedia tools
6. Change in teaching methodology brings about dissatisfaction since most teachers prefer the traditional methods for teaching
7. Less privileged lecturers (lecturers with disability) inability to access SNT system
8. Higher cost of licensing the SNTs and purchase of sophisticated devices
9. Exposure to chemicals leading to health problems

The mean values obtained from the table 1 above compared to the mean value of 3.50 (the decision mean value), show that all respondents agree to the fact that SNT creates opportunities for lecturers to improve their teaching methodology and improve student involvement.

From the table 1 above, it could be realized that majority of the respondents considered the following challenges as the most prevalent among the listed items. This is noticeable with reference to the mean values obtained per the survey results. These are ranked with the most challenging factor first.

- Higher cost of licensing the SNTs and purchase of sophisticated devices (mean value = 5.35)
- Incompetence in the use of multimedia tools (mean value = 5.31)
- Safety and privacy issues (mean value = 5.21)
- Change in teaching methodology brings about dissatisfaction since most teachers prefer the traditional methods for teaching (mean value = 4.69)
- The notion of uncertainty due to lack of exposure to SNTs (mean value = 4.48)

Lecturers in these private/public universities consider cost as the most challenging factor when it comes to the implementation of SNTs in teaching especially in a developing country like Ghana.

The findings are also in line and support the findings of Zaidieh(2012) in which the author stated challenges such as privacy, taking time up and miscommunication to hinder the implementation of SNTs. The findings of Education MS(2010) report, educators who have used social networking technology are more positive about the benefits as compared to those who have not. This supports the notion of uncertainty due to lack of exposure to SNTs as listed as part of the challenges in this research.

The respondents also supported the idea of misuse of time. This is because the respondents argued that there are no clear systems in put in place for monitoring the activities of learners on a regular basis. This is a barrier to the implementation of social networking in schools. Lack of monitoring mechanism to check user activity makes it a challenge to implement SNTs in teaching.

Similarly, the findings agree with research reports of Hamid (2011), that time management is an issue lecturer’s face when it comes to the implementation of online social networks for teaching. They also reported or argued that the challenges of using social networks in higher education could be lecturer-specific. This is because their findings showed that some lecturers had not faced any major problems in using the social technologies.

This observation might be true especially in the case of proactive lecturers who are appropriating and using these technologies without being asked to do so by the faculty. On the contrary, others argued that it is challenging to introduce students to a new tool that was not familiar to them.

According to Hamid (2011), implementing social network technologies is a challenge since most students have limited access to the Internet. Additionally, some students did not have the latest range of mobile phones that are capable of running mobile social technologies, particularly for Twitter and Facebook.

Some students could not afford to own smart phones although all of them carried mobile phones. Due to this reason also, the use of social technologies could only be considered as complementary due to limitation in its ubiquitous access. This also agrees with the point made earlier that there is a high cost of licensing SNTs and

purchasing other sophisticated devices to support the usage of social networks in teaching.
In Ghana, most teachers do not welcome the new technologies that aid in teaching. Change in teaching methodology brings about dissatisfaction since most teachers prefer the traditional methods for teaching. This is consistent with the report by Hu (2005) that traditional means of teaching and instruction still dominate, especially in developing countries.

This then becomes a challenge introducing social network technologies in teaching in private universities in Ghana.

7.2 Findings on research question 2

The findings from research question 2 indicated in table 2 show that there is great deal of opportunities for lecturers who adopt these technologies in their teaching. In the field of education, social-networking sites offer a student the opportunity to connect with other students, educators, administrators, alumni, both within and outside his current institution (Zamza, n.d.).

This offer is not limited to students alone but, the faculty as well since it creates the platform for lecturers to communicate with their students, post quizzes and assignments and other relevant materials to students. The faculty members get the opportunity to connect with other colleagues within the institution and without.

Scholars praise social-networking tools for their capability to attract, motivate and engage students in meaningful communicative practice, content exchange, and collaboration (Mills, 2012).

To emphasize again, despite the challenges, its advantages or opportunities cannot be overlooked. Lecturers must capitalize on the numerous opportunities and implement the social network tool in teaching. Some of these opportunities are as follows:

- It supports different teaching methods/styles due to its flexibility
- It creates effective teaching and learning environment
- Teaching location is independent
- It brings up creativity and innovation
- It is convenient and easy to use for teaching due to its interactivity
- It enhances quality academic outcomes due to its numerous opportunities

The research findings are consistent with the finds of Dooley (1999) in which they discussed that in a teaching environment where SNTs are implemented or adopted for lecturing, as the lecture progresses, a student may miss something or maybe unclear on a certain point (in the traditional classroom).

The learning environment addresses this problem in a social media centric way. It allows the student to ask the question not just of the instructor but, of his fellow classmates. This crowd-sourcing allows the student to catch up or gain clarification without disrupting the entire class or limiting his answer pool to those classmates directly adjacent to him, and it allows him to do so in a minimally disruptive way.

This point supports the point made in this research that it is convenient and easy to use due to its interactivity and creates effective teaching and learning environment.
The findings are consistent with the report by Brown (2010) and Schroeder et al. (2010) which states that the driving factors for adoption of online social networks include the increasingly ubiquitous access, ease of use, functionality, and flexibility of social technologies.

8. Conclusions

There are some implementation challenges and opportunities that face the use of social networking technologies as educational tool in private tertiary institutions in Ghana. The key indicators or factors dominating among the challenges are higher cost of licensing the SNTs and purchase of sophisticated devices, Incompetence in the use of multimedia tools, safety and privacy issues, change in teaching methodology brings about dissatisfaction since most teachers prefer the traditional methods for teaching and the notion of uncertainty due to lack of exposure to SNTs are the major challenges that face the implementation of social network technologies in private universities.

Other factors such as exposure to chemicals leading to health problems such as eye related problems, back ache, wrist injuries and the like and the inability for lecturers with disability to use the technology are also challenges that affects the implementation of SNTs in teaching.

On the other hand, supporting different teaching methods due to its flexibility, creating effective teaching and learning environment, independent teaching location, convenience and ease of use and interactivity and enhancement of academic performance have created vast opportunities and have vital influence in the use of social networking in education (for teaching and learning).

The research paper addressed some of previous studies about the challenges and opportunities in the implantation of SNTs for teaching and learning in education especially distance education and other online programs.

9. Recommendation

Information Communication Technologies offers immense opportunity for both private and public universities in developing countries to advance their teachings and learning processes to equip both faculty and students to compete with colleagues in the developed countries.

The use ICT tools are gradually closing the gap between educators and students. So far, most of the universities in developing countries possess basic ICT infrastructure such as Local Area Network (LAN), internet, computers, video, audio, CDs and DVDs, and mobile technology facilities that form the basis for the establishment of e-learning. It is argued that, universities in developing countries should adopt social network technologies to improve teaching and learning processes. Therefore, the university management should consider integrating ICT tools such as SNTs in teaching and learning practices.

Management needs to train its faculty on how to use cutting-edge technologies such as SNTs to be implemented in teaching. They should regularly organize seminars and workshops that will enhance the knowledge base of these faculties on the new trends of technology and students on how to use the SNTs for teaching and learning respectively. They should include a comprehensive budget covering the maintenance of the entire infrastructure for running the SNTs.

This will help since SNTs require payment to use some features, like video conferencing and the like to be enabled. A good wireless network should be situated in every private university. Access to network only through cables would affect mobility of the lecturers and the students in the usage of the technology. At worse case, hot spots should be created on campus where students can cluster around and use.
University management should consider how to make laptops available for all of their lecturers.

When lecturers have laptops that will encourage them to have a strong affiliation with the technology and they will overcome the lack of exposure and change their teaching methods and move away from the traditional methods of the teaching. It will also boost their frequency of usage. The absence of the laptops means that lecturers will still be computer phobia and will resist the technology despite, its vast benefit.

References

[1] E. Turban, D. King, J. Lee, & D. Viehland, “E-commerce A managerial Perspective New Jersey 07458”, Pearson Education, Inc., Upper Saddle River, 2010, pp. 392.
[2] S. Hamid, J. Waycott, S. Chang, S. Kurnia, “Appropriating Online Social Networking (OSN) Activities for Higher Education: Two Malaysian Cases”, Proceedings ascilite, 2011, pp. 526 – 537.
[3] D. Lee, “Web based instruction in China: Pedagogical implication”, ETR&D 52(1), 2004.
[4] W. Guo, & Y. Cai, “Responses of Chinese Higher Education to the Information Society”, E-Learning, vol. 3(1), 2006, pp. 353-360.
[5] C. Redecker, K. Ala-Mutka, & Y. Punie, “Learning 2.0- The impact of Social Media on Learning in Europe”, JRC Technical Notes
[6] J. Picardo, “Teaching and Learning with Social Networks: Barriers to Adoption”, Network.Ed, 2011 [Date retrieved from http://www.josepicardo.com/2011/08/teaching-and-learning-with-social-networks-barriers-to-adoption/ on 20th August, 2012]
[7] K.E Dooley, “Towards a Holistic Model for the Diffusion of Educational Technologies: An Integrative Review of Educational Innovation Studies”, Educational Technology & Society. ISSN 1436-452. [Date Retrieved from http://www.ifets.info/journals/2_4/kim_dooley.html on 21st August, 2012]
[8] G. Hordemann & J. Chao, “Design and Implementation Challenges to an Interactive Social Media Based Learning Environment”, Interdisciplinary Journal of Information, Knowledge, and Management 7 (1) Special Section on Social Networking, Teaching, and Learning, 2012, pp. 92-94
[9] J.J. Hu, “Information and communication technology in education in China: Policies and practices”, A seminar paper presented at Hong Kong, 2005. [Date Retrieved from: http://hdl.handle.net/10722/43997[accessed 21February 2013]
[10] M.S Education, “School Principals and Social Networking in Education: Practices, Policies, and Realities”, 2010, pp. 1-10.
[11] N. Mills, “Situated Learning through Social Networking Communities: The Development of Joint Enterprise, Mutual Engagement, and a Shared Repertoire”, CALICO Journal, vol. 28(2), 2011, pp. 345-368.
[12] S. Brown, “From VLEs to learning webs: the implications of Web 2.0 for learning and teaching”, Interactive Learning Environments, vol. 18(1), 2010, pp. 1-10.
[13] A. Schroeder, S. Minocha, & C. Schneider, “Social Software in Higher Education: The Diversity of Applications and Their Contributions to Students’ Learning Experiences”, Communications of the Association for Information Systems, Vol. 25(1), 2010, pp. 547-564.
[14] I. Jung, “ICT and quality assurance to support ubiquitous access to distance education: Promises, realities and recent

breakthroughs." 5th EDEN Research Workshop - slideshare.net/eden_online, 20th-30th Oct 2008

[15] I. N. Zamzam "Challenges and Opportunities of Technology Based Instruction in Open and Distance Learning: A Comparative Study of Tanzania and China." Faculty of Education, the Open University of Tanzania, Retrieved from http://www.ubuntunet.net/sites/ubuntunet.net/files/uc2012_presentationzamzam.pd on February 31, 2013]

[16] J. Y. Zaidieh, "The use of social networking in Education: Challenges and Opportunities", World of Computer Science and Information Technology Journal (WCSIT) ISSN: 2221-0741, Vol. 2, No. 1, 2012, pp. 18-21.

[17] W.D. Chapman, A. Garrett, & O.L. Mahlck, "Adapting technology for school improvement: A global perspective." International Institute of Educational Planning, 2004

[18] A.S. Sife, E.T Lwoga, & C. Sanga, "New technologies for teaching and learning: Challenges for higher learning institutions in developing countries", International Journal of Education and Development using Information and Communication Technology (IJEDICT), Vol. 3, Issue 2, 2007, pp. 57-67.

Mr. Kofi Sarpong Adu-Manu holds Bachelor's and Master's degree in Computer Science and works with the Department of Computer at Valley View University, Accra-Ghana. I have taught courses in major areas in Computer Science and Information Technology, including Programming, Distributed Systems, Computer Graphics, Visualization and Simulation, Database systems, and Multimedia at Valley View University. I serve as a Senior Research Assistant at the Faculty of Science and have supervised final year projects in software development. As part of the author's academic achievements, Adu-Manu was won three (3) best paper awards and has seven (7) published papers to his credit. My research focus is on Data warehousing, programming languages, Computer Graphics and Social Networking. I serve as a board member to the Clute Institute in Colorado, USA and a member of Computer Science Teachers Association (CSTA) in the United States of America.

Mr. Arthur John Kingsley a faculty member at the institute of Computer Science of Valley View University. I hold Bachelor of Science in Computer Science and Master in Information Technology. I have keen interest computer Science and Information Technology research areas of Educational Technologies, Web and Mobile applications. I have professional teaching membership with Victoria Information Technology Teachers Association (VITTA) in Australia and Computer Science Teachers Association (CSTA) in USA

Mr. Clement Yeboah holds BBA in Accounting and MBA in Accounting. At the moment I am pursuing MBA in Finance at Strayer University, Columbus campus - USA.

BER Analysis of 3x3 MIMO Spatial Multiplexing under AWGN & Rician Channels for Different Modulation Techniques

Anuj Vadhera[1], Lavish Kansal[2]

[1] School of Electronics Engineering, Lovely Professional University
Phagwara, Punjab, India

[2] Department of Electronics and Communication, Lovely Professional University
Phagwara, Punjab, India

Abstract

The multiple input multiple output (MIMO) antenna system provides very promising gain in capacity without increasing the use of spectrum, throughput, and power consumption. This is also less sensitive to fading, hence leading to a breakthrough in the data rate of wireless communication systems. In such systems, the antenna properties as well as the multipath channel characteristics play a key role in determining communication performance.

This paper proposes the analysis and performance of Spatial Multiplexing technique of MIMO system.Here different fading channels like AWGN and Rician are used for analysis purpose. Moreover we analyzed the technique using high level modulations (i.e. M-PSK for different values of M). Detection algorithms used are Zero- Forcing and Minimum mean square estimator.

Keywords: *Multiple Input Multiple Output (MIMO), Spatial Multiplexing(SM), Additive White Gaussian Noise (AWGN), Zero-Forcing (ZF) , Minimum Mean Square Estimator (MMSE).*

1. Introduction

The idea of using multiple receive and multiple transmit antennas has emerged as one of the most significant technical breakthroughs in modern wireless communications. As a result, MIMO is considered a key technology for improving the throughput of future wireless broadband data systems. MIMO is the use of multiple antennas at both the transmitter and receiver to improve communication performance. It is one of several forms of smart antenna technology. A core idea in MIMO system is space–time signal processing in which time (the natural dimension of digital communication data) is complemented with the spatial dimension inherent in the use of multiple spatially distributed antennas. A key feature of MIMO systems is the ability to turn multipath propagation, traditionally a pitfall of wireless transmission, into a benefit for the user.

Two typical approaches in the MIMO systems are to provide diversity gain as in space-time coding (STC) or to allow spatial multiplexing (SM). While STC systems are capable of improving system reliability through coding across space and/or time, SM systems are capable of increasing data transmission rate through spatial multiplexing. In this paper, we focus on the SM technique.

Spatial multiplexing is a transmission technique in MIMO wireless communication to transmit independent and separately encoded data signals, so-called streams, from each of the multiple transmit antennas. Therefore, the space dimension is reused, or multiplexed, more than one time. If the transmitter is equipped with Nt antennas and the receiver has Nr antennas, the maximum spatial multiplexing order (the number of streams) is

$$Ns = \min(Nt, Nr) \qquad (1)$$

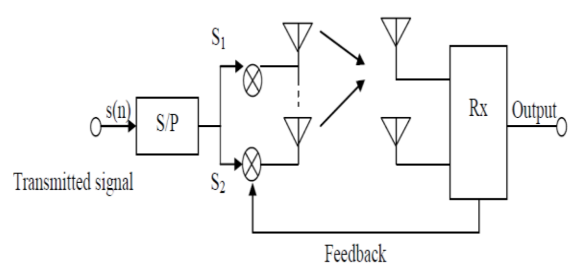


Fig. 1 Spatial Multiplexing Concept

MIMO spatial multiplexing achieves this by utilizing the multiple paths and effectively using them as additional channels to carry data such that receiver receives multiple data at the same time. The tenet in spatial multiplexing is to transmit different symbols from each antenna and the receiver discriminates these symbols by taking advantage of the fact that, due to spatial selectivity, each transmit antenna has a different spatial signature at the receiver. This allows an increased number of information symbols per MIMO symbol. In any case for MIMO spatial multiplexing, the number of receiving antennas must be equal to or greater than the number of transmit antennas such that data can be transmitted over different antennas. Therefore the space dimension is reused or multiplexed more than one time. The data streams can be separated by equalizers if the fading processes of the spatial channels are nearly independent. Spatial multiplexing requires no bandwidth expansion and provides additional data bandwidth in multipath radio scenarios [2].

The general concept of spatial multiplexing can be understood using MIMO antenna configuration. In spatial multiplexing, a high data rate signal is split into multiple lower data rate streams and each stream is transmitted from a different transmitting antenna in the same frequency channel. If these signals arrive at the receiver antenna array with different spatial signatures, the receiver can separate these streams into parallel channels thus improving the capacity. Thus spatial multiplexing is a very powerful technique for increasing channel capacity at higher SNR values. The maximum number of spatial streams is limited by the lesser number of antennas at the transmitter or receiver side. Spatial multiplexing can be used with or without transmit channel knowledge.

2. Multiple Input Multiple Output (MIMO)

Multiple antennas can be used at the transmitter and receiver, an arrangement called a MIMO system. A MIMO system takes advantage of the spatial diversity that is obtained by spatially separated antennas in a dense multipath scattering environment. MIMO systems may be implemented in a number of different ways to obtain either a diversity gain to combat signal fading or to obtain a capacity gain. Generally, there are three categories of MIMO techniques. The first aims to improve the power efficiency by maximizing spatial diversity. Such techniques include delay diversity, STBC and STTC. The second class uses a layered approach to increase capacity. One popular example of such a system is V-BLAST suggested by Foschini et al. [2] where full spatial diversity is usually not achieved. Finally, the third type exploits the knowledge of channel at the transmitter. It decomposes the channel coefficient matrix using SVD and uses these decomposed unitary matrices as pre- and post-filters at the transmitter and the receiver to achieve near capacity [3].

2.1 Benefits of MIMO system

MIMO channels provide a number of advantages over conventional Single Input Single Output (SISO) channels such as the array gain, the diversity gain, and the multiplexing gain. While the array and diversity gains are not exclusive of MIMO channels and also exist in single-input multiple-output (SIMO) and multiple-input single-output (MISO) channels, the multiplexing gain is a unique characteristic of MIMO channels. These gains are described in brief below

Array Gain

Array gain denotes the improvement in receive signal-to-noise ratio (SNR) that results from a coherent combining effect of the information signals. The coherent combining may be realized through spatial processing at the receive antenna array and/or spatial pre-processing at the transmit antenna array. Formally, the array gain characterizes the horizontal shift of the error probability versus transmitted or received power curve (in a log-log scale), due to the gain in SNR

Spatial Diversity Gain

Diversity gain is the improvement in link reliability obtained by receiving replicas of the information signal through (ideally independent) fading links. With an increasing number of independent copies, the probability that at least one of the signals is not experiencing a deep fade increases, thereby improving the quality and reliability of reception. A MIMO channel with nT transmit and Nr receive antennas offers potentially nTnR independently fading links and, hence, a spatial diversity order of nTnR. Formally, the diversity gain characterizes the slope of the error probability versus transmitted or received power curve (in a log-log scale) in the high-SNR regime.

Spatial Multiplexing Gain

MIMO systems offer a linear increase in data rate through spatial multiplexing, i.e., transmitting multiple, independent data streams within the bandwidth of operation. Under suitable channel conditions, such as rich scattering in the environment, the receiver can separate the data streams. Furthermore, each data stream experiences at least the same channel quality that would be experienced by a SISO system, effectively enhancing the capacity by a multiplicative factor equal to the number of substreams. In general, the number of data streams that can be reliably supported by a MIMO channel coincides with the minimum of the number of transmit antennas T and the number of receive antennas nR, i.e., min{nT; nR}.

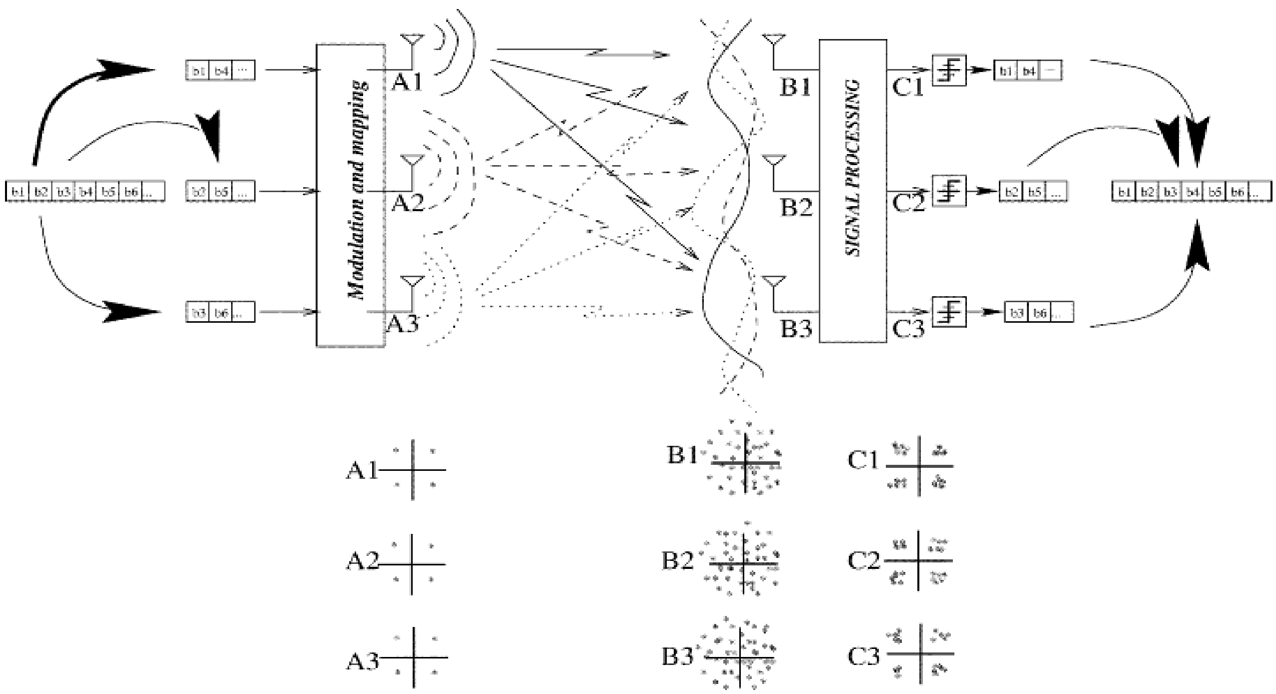


Fig. 2 Basic spatial multiplexing (SM) scheme with three Transmit and three Receive antennas Ai, Bi, and Ci represent symbol constellations for the three inputs at the various stages of transmission and reception.

2.2 Channels

Wireless transmission uses air or space for its transmission medium. The radio propagation is not as smooth as in wire transmission since the received signal is not only coming directly from the transmitter, but the combination of reflected, diffracted, and scattered copies of the transmitted signal.

2.2.1 AWGN Channel

Additive white Gaussian noise (AWGN) channel is universal channel model for analyzing modulation schemes. In this model, the channel does nothing but add a white Gaussian noise to the signal passing through it. This implies that the channel's amplitude frequency response is flat (thus with unlimited or infinite bandwidth) and phase frequency response is linear for all frequencies so that modulated signals pass through it without any amplitude loss and phase distortion of frequency components. Fading does not exist. The only distortion is introduced by the AWGN. AWGN channel is a theoretical channel used for analysis purpose only. The received signal is simplified to:

$$r(t)=x(t)+n(t) \tag{2}$$

where x(t) is the transmitted signal and n(t) is additive white Gaussian noise.

2.2.1Rician Channel

When there is line of sight, direct path is normally the strongest component goes into deeper fade compared to the multipath components. This kind of signal is approximated by Rician distribution. As the dominating component run into more fade the signal characteristic goes from Rician to Rayleigh distribution. The received signal can be simplified to:

$$r(t)=x(t)*h(t)+n(t) \tag{3}$$

where h(t) is the random channel matrix having Rician distribution and n(t) is the additive white Gaussian noise. The Rician distribution is given by:

$$P(r) = \frac{r^2}{\sigma^2} e^{(-\frac{r^2+A^2}{\sigma^2})} I_0\left(\frac{A_r}{\sigma^2}\right) \quad \text{for } (A \geq 0 , r \geq 0) \tag{4}$$

where A denotes the peak amplitude of the dominant signal and I0[.] is the modified Bessel function of the first kind and zero-order.

2.3 Modulation

Modulation is the process of mapping the digital information to analog form so it can be transmitted over the channel. Consequently every digital communication system has a modulator that performs this task. Closely related to modulation is the inverse process, called demodulation, done by the receiver to recover the transmitted digital information [4].

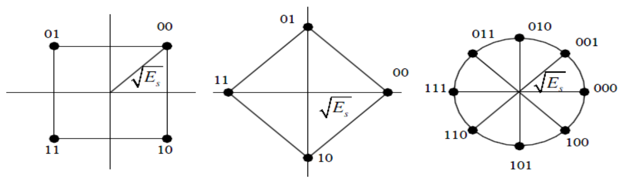
Modulation of a signal changes binary bits into an analog waveform. Modulation can be done by changing the amplitude, phase, and frequency of a sinusoidal carrier. There is several digital modulation techniques used for data transmission.

2.3.1 Phase Shift Keying

Phase Shift Keying is a digital modulation scheme that conveys data by changing or modulating, the phase of a reference signal (the carrier wave). In M-ary PSK modulation, the amplitude of the transmitted signals was constrained to remain constant, thereby yielding a circular constellation. Modulation equation of M-PSK signal is:

$$s_i(t) = \sqrt{\frac{2E_S}{T}}\cos\left(2\pi f_c t + \frac{2\pi i}{M}\right) \qquad (5)$$

i=0,1….,M



(a) (b) (c)
Fig. 3 Constellation Diagrams of M-PSK
a) QPSK (b) QPSK (c) 8-PSK

2.4 Detection Techniques

There are numerous detection techniques available with combination of linear and non-linear detectors. The most common detection techniques are ZF, MMSE and ML detection technique. The generalized block diagram of MIMO detection technique is shown in Figure 4.

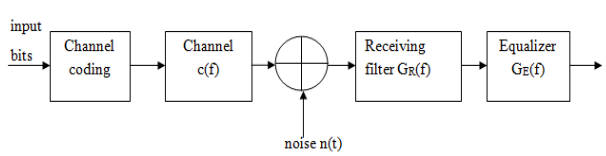


Fig. 4 Block Diagram of system with equalizer

2.4.1 Zero Forcing detection

The ZF is a linear estimation technique, which inverse the frequency response of received signal, the inverse is taken for the restoration of signal after the channel. The estimation of strongest transmitted signal is obtained by nulling out the weaker transmit signal. The strongest signal has been subtracted from received signal and proceeds to decode strong signal from the remaining transmitted signal. ZF equalizer ignores the additive noise and may significantly amplify noise for channel.

The basic Zero force equalizer of 2x2 MIMO channel can be modeled by taking received signal y_1 during first slot at receiver antenna as:

$$r_1 = h_{1,1}x_1 + h_{1,2}x_2 + n_1 = [h_{1,1} \quad h_{1,2}]\begin{bmatrix}x_1\\ x_2\end{bmatrix} + n_1 \qquad (6)$$

The received signal y_2 at the second slot receiver antenna is:

$$r_2 = h_{2,1}x_1 + h_{2,2}x_2 + n_2 = [h_{2,1} \quad h_{2,2}]\begin{bmatrix}x_1\\ x_2\end{bmatrix} + n_2 \qquad (7)$$

Where i=1, 2 in x_i is the transmitted symbol and i=1, 2 in $h_{i,\ j}$ is correlated matrix of fading channel, with j represented transmitted antenna and i represented receiver antenna, is the noise of first and second receiver antenna. The ZF equalizer is given by:

$$W_{ZF} = (H^H)^{-1}H^H \qquad (8)$$

Where W_{ZF} is equalization matrix and H is a channel matrix. Assuming $M_R \geq M_T$ and H has full rank, the result of ZF equalization before quantization is written as:

$$y_{ZF} = (H^H H)^{-1}H^H y \qquad (9)$$

2.4.2. Minimum Mean Square Estimator (MMSE)

Minimum mean square error equalizer minimizes the mean –square error between the output of the equalizer and the transmitted symbol, which is a stochastic gradient algorithm with low complexity. Unlike a ZF equalizer, an MMSE equalizer maximizes the signal–to distortion ratio by penalizing both residual ISI and noise enhancement. Instead of removing ISI completely, an MMSE equalizer allows some residual ISI to minimize the overall distortion. Compared with a ZF equalizer, an MMSE equalizer is much more robust in presence of deepest channel nulls and noise. Most of the finite tap equalizers are designed to minimize the mean square error performance metric but MMSE directly minimizes the bit error rate. The channel model for MMSE is same as ZF [13],[14]. The MMSE equalization is

$$W_{MMSE} = arg_G^{min} E_{x,n}[\|x - x^{\wedge}\|^2] \qquad (10)$$

Where is W_{MMSE} equalization matrix, H channel correlated matrix and n is channel noise

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 1, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

$$y_{MMSE} = H^H(HH^H + n_o I_n)^{-1} y \qquad (11)$$

3. Results and Discussions

The system discussed above has been designed and results are shown in the form of SNR vs. BER plot for different modulations and different channels. Here antenna configuration 3x3 is analysed using ZF and MMSE detection techniques. Analyses have been done for two channels AWGN and Rician channel using MMSE and ZF detection algorithms.

3.1 Using ZF detection algorithm:

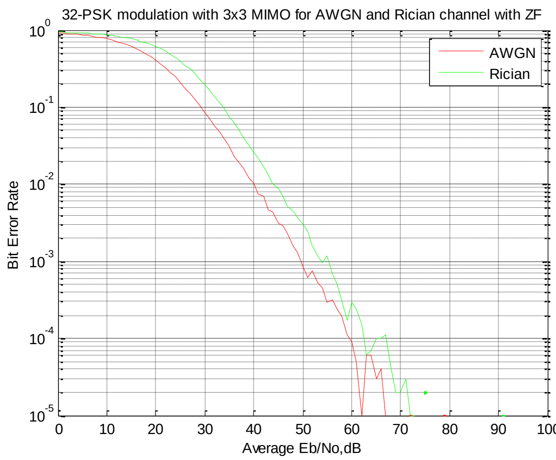


Fig. 5(a)

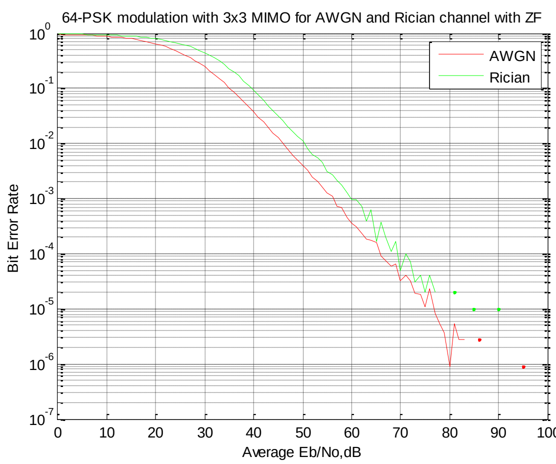


Fig 5(b)

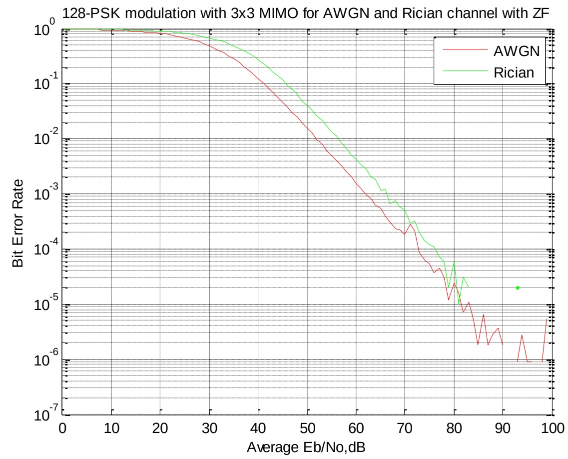


Fig. 5(c)

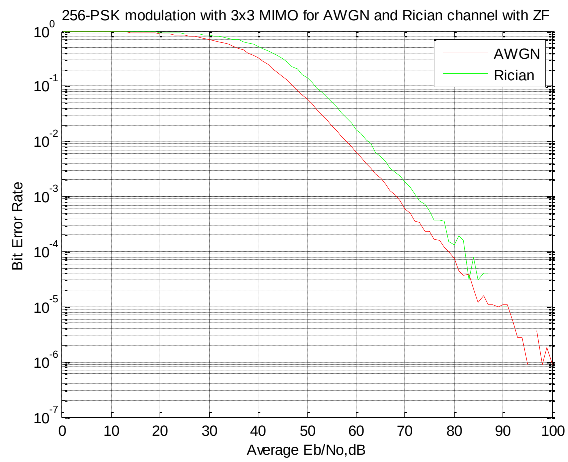


Fig. 5(d)

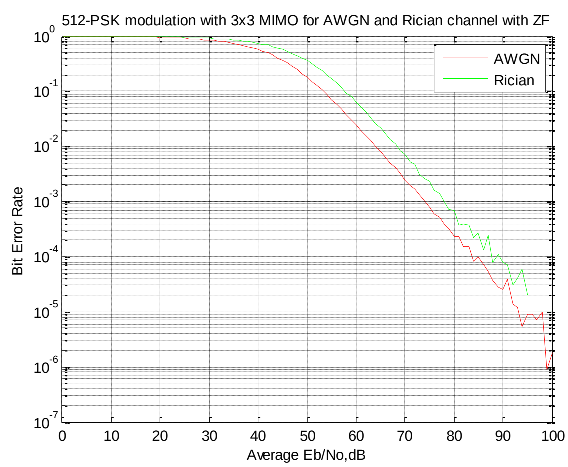


Fig. 5(e)

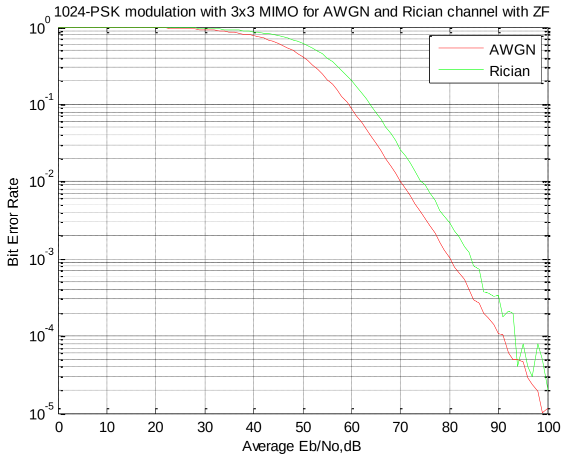


Fig. 5(f)

Fig. 5 BER vs. SNR plots over AWGN & Rician channel for SM technique using 3x3 MIMO using ZF Equalization a) 32 PSK b) 64 PSK c) 128 PSK d) 256 PSK e) 512 PSK f) 1024 PSK

Table 1. Comparison of different Modulation Techniques for AWGN & Rician Channel for 3x3 MIMO Spatial Multiplexing using ZF Equalization

Modulations	Rician Channel	AWGN Channel	Improvement
32-PSK	62dB	59 dB	3 dB
64-PSK	70dB	66 dB	4 dB
128-PSK	77dB	73 dB	4 dB
256-PSK	82 dB	79 dB	3 dB
512-PSK	88 dB	84 dB	4 dB
1024-PSK	93 dB	90 dB	3 dB

Table 1 presents that at 32-PSK, 256-PSK, 1024-PSK there is an improvement of 3dB and at 64-PSK, 128-PSK and 512-PSK there is an improvement of 4dB at BER 10^{-4}. Hence table 1. shows the improvement in terms of decibels shown by proposed system employing SM technique for 3x3 MIMO system for different modulation schemes over different environments (channels).

3.2 Using MMSE detection algorithm:

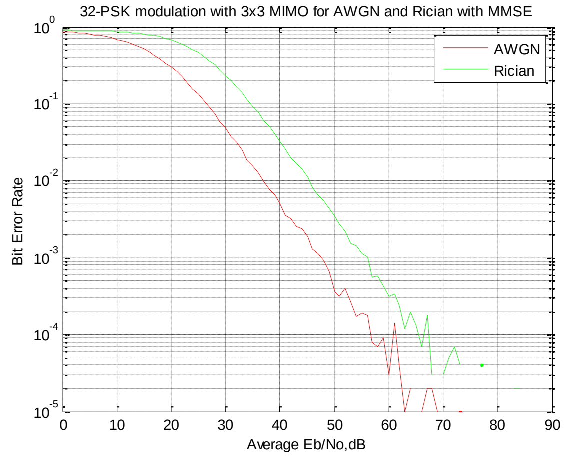


Fig. 6(a)

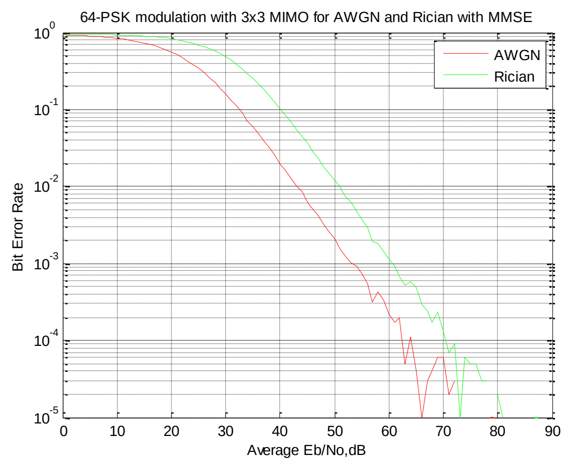


Fig. 6(b)

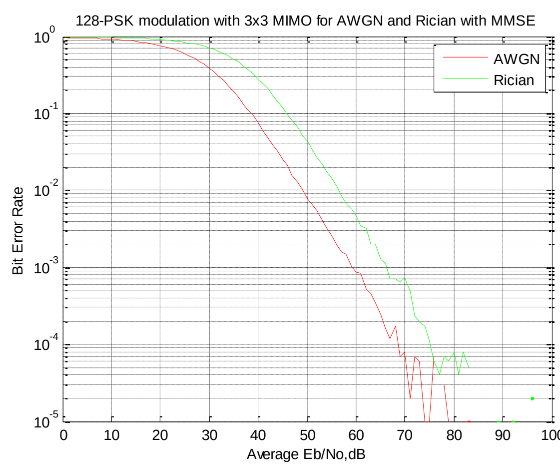


Fig. 6(c)

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 5, No 1, September 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

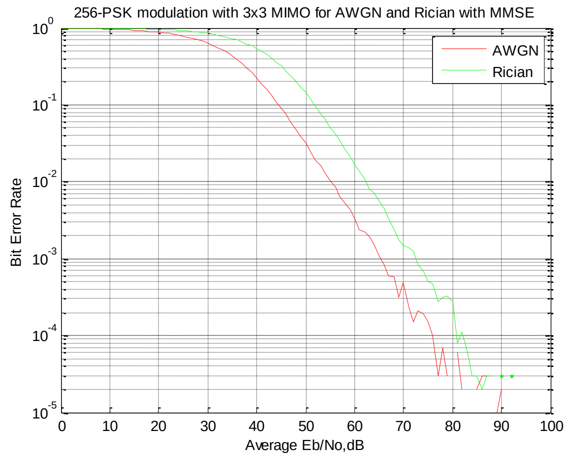


Fig. 6(d)

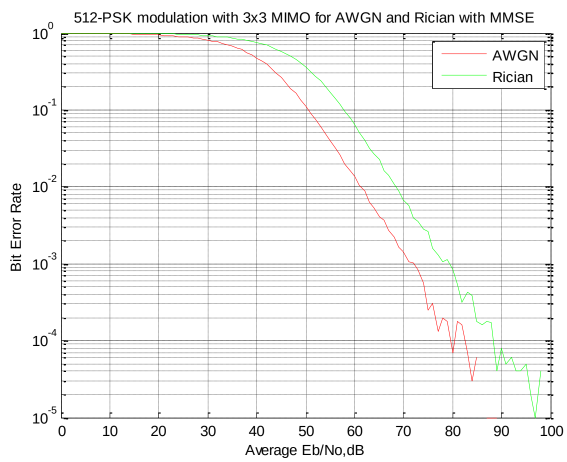


Fig. 6(e)

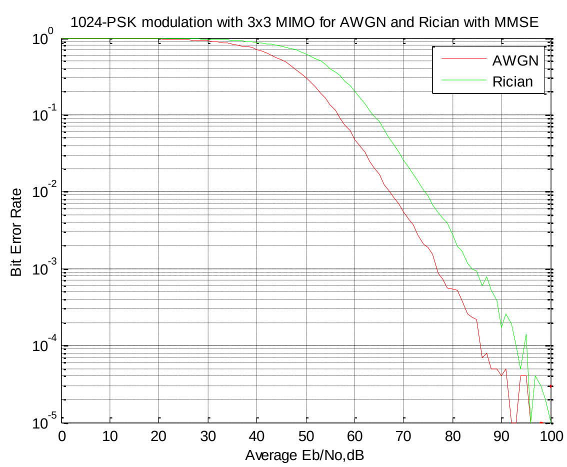


Fig. 6(f)

Fig. 6 BER vs. SNR plots over AWGN & Rician channel for SM technique using 3x3 MIMO using MMSE Equalization a) 32 PSK b) 64 PSK c) 128 PSK d) 256 PSK e) 512 PSK f) 1024 PSK

Table 2. Comparison of different Modulation Techniques for AWGN & Rician Channel for 3x3 MIMO Spatial Multiplexing using MMSE Equalization

Modulations	Rician Channel	AWGN Channel	Improvement
32-PSK	65dB	62 dB	3 dB
64-PSK	70 dB	64 dB	6 dB
128-PSK	75 dB	69 dB	6 dB
256-PSK	80 dB	71 dB	9 dB
512-PSK	85 dB	77 dB	8 dB
1024-PSK	95 dB	86 dB	9 dB

Table 2 depicts that at 64-PSK, 128-PSK there is an improvement of 6dB, at 256-PSK, 1024-PSK there is an improvement of 9dB and at 512-PSK there is an improvement of 8dB.Hence table 2 shows improvement in terms of decibels shown by proposed system employing SM technique for 3x3 MIMO system for different modulation schemes over different environments (channels) at BER 10^{-4}.

4. Conclusions

In this paper, an idea about the performance of the MIMO-SM technique at higher modulation levels and for 3x3 antenna configuration using different signal detection techniques is presented. MIMO-SM technique can be implemented using higher order modulations to achieve large data capacity. But there is a problem of BER (bit error rate) which increases as the order of the modulation increases. The solution to this problem is to increase the value of the SNR so, that the effect of the distortions introduced by the channel will also goes on decreasing, as a result of this, the BER will also decreases at higher values of the SNR for high order modulations.

Several different diversity modes are used to make radio communications more robust, even with varying channels. These include time diversity (different timeslots and channel coding), frequency diversity (different channels, spread spectrum, and OFDM), and also spatial diversity. Spatial diversity requires the use of multiple antennas at the transmitter or the receiver end. Multiple antenna systems are typically known as Multiple Input, Multiple Output systems (MIMO). Multiple antenna technology can also be used to increase the data rate (spatial multiplexing) instead of improving robustness. In future, we can make a single integrated circuit that uses both methods combination.

Acknowledgments

With deep sense of gratitude, I express my sincere thanks to my esteemed and worthy guide Mr. Lavish Kansal, Assistant Professor, Electronics and Communication Engineering Department, Lovely Professional University, Phagwara, for his valuable advice, motivation, guidance, encouragement, efforts, timely help and the attitude with which he solved all of my queries regarding thesis work.

References

[1] H. Jiang and P. A. Wilford, "A hierarchical modulation for upgrading digital broadcasting systems," IEEE Transaction on Broadcasting, vol. 51, pp. 222-229, June 2005.
[2] P. W. Wolniansky, G. J. Foschini, G. D. Golden and R. A.Valenzuela, "V-BLAST: an architecture for realizing very high data rates over the rich- scattering wireless channel," In Proceeding of International symposium on Signals, Systems Electronics, pp. 259-300, October 1998.
[3] J. Ha, A. N. Mody, J. H. Sung, J. Barry, S. Mclaughlin and G. L. Stuber, "LDPC coded OFDM with Alamouti/SVD diversity technique," IEEE Journal on Wireless Personal Communication, Vol. 23 , Issue 1,pp. 183-194,Oct. 2002.
[4] P. S. Mundra , T. L. Singal and R. Kapur, "The Choice of A Digital Modulation ,Schemes in A Mobile Radio System", In proceedings of IEEE Vehicular Technology Conference, Issue 5, pp 1-4,(Secaucus, NJ)1993.
[5] P. Liu & I1-Min Kim, "Exact and Closed-Form Error Performance Analysis for Hard MMSE-SIC Detection in MIMO Systems", IEEE Transactions on Communication, Vol. 59, no. 9, September 2011.
[6] P. Sanghoi & L. Kansal," Analysis of WiMAX Physical Layer Using Spatial Multiplexing Under Different Fading Channels", SPIJ, Vol.(6),Issue(3),2012.
[7] C. Wang & E. K. S. Au, R. D Murch, W. H. Mow & V. Lau," On the Performance of the MIMO Zero-Forcing Receiver in the Presence of Channel Estimation Error", IEEE Transactions on Wireless Communication, Vol. 6,no.3,2007.
[8] X. Zhang, Y. Su & G. Tao, "Signal Detection Technology Research of MIMO-OFDM System", 3rd International Congress on Image and Signal Processing, pp 3031-3034, 2010.
[9] I. Ammu & R. Deepa, "Performance Analysis of Decoding Algorithms in multiple antenna systems", IEEE, pp 258-262, 2011.
[10] H. B. Voelcker, "Phase-shift keying in fading channels" , In IEEE Proceeding on Electronics and Communication Engineering, Vol. 107, Issue 31, pp 31-38, 1960.
[11] D. S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multi-element antenna systems", IEEE Transaction on Communication, Vol. 48, pp. 502–513, 2000.
[12] G. J. Foschini, K. Karakayali, and R. A.Valenzuela, "Coordinating multiple antenna cellular networks to achieve enormous spectral efficiency," Communications, IEE Proceedings, Vol. 153, pp. 548–555, 2006.
[13] J. S. Thompson, B. Mulgrew and Peter M. Grant, "A comparison of the MMSE detector and its BLAST versions for MIMO channels", IET seminar on Communication System from Concept to Implementation, pp. 1911-1916, 2001.
[14] X. Zhang and Sun-Yuan Kung, "Capacity analysis for parallel and sequential MIMO equalizers", IEEE Transaction on Signal processing, Vol. 51, pp. 2989-3002, 2003.

Anuj Vadhera was born in Fazilka. She received her B.Tech degree in Electronics and Communication Engineering from Punjab Technical University, Jalandhar, in 2009, and pursuing M.Tech degree in Electronics and communication engineering from Lovely Professional University, Phagwara, India. Her research interests include MIMO systems, cognitive radios and wireless systems.

Lavish Kansal was born in Bathinda. He received his B.Tech degree in Electronics and Communication Engineering from PTU, Jalandhar in 2009 and M.E. degree in Electronics and Communication Engineering from Thapar University, Patiala in 2011.He is working as Assistant Professor in the department of Electronics and communication Engineering, Lovely Professional University, Phagwara, India. He has published 15 papers in International journals. His research area includes Digital Signal Processing, Digital Communication & Wireless Communication.

Network research based on the fuzzy comprehensive evaluation model of natural language

Zhi-hong Ma[1] ， Xun-song He[2]，Hao-xuan Ding[3]

**1 Basic Science Department，Tianjin Agricultural University
Tianjin 300384 China**

**2 Department of Computer Science and Information Engineering，Tianjin Agricultural University
Tianjin, 300384 , China**

**3 Department of Computer Science and Information Engineering，Tianjin Agricultural University
Tianjin, 300384 , China**

Abstract

The size of the possibility of determining the criminal conspiracy helps to survey, monitor or question the most likely suspects. However , we can make some unclear boundary and factors that are not easy to quantitative quantified by using the fuzzy comprehensive evaluation of principle. In this paper, the quantification of the theme of the dialogue of the network crime gang draw a priority list of a criminal conspiracy. Compared with the semantics of message transmission analysis and text analysis, the fuzzy comprehensive evaluation of principle not only makes the theme for the conspiracy more authentic intuitively and improves the efficiency of the infiltration of the core of the criminal gang's conspiracy.

Keywords: *Criminal conspiracy , Fuzzy, comprehensive evaluation , Natural language*

1. Introduction

The current case has 83 nodes, 400 links (some involving more than 1 topic), over 21,000 words of message traffic, 5 topics (3 have been deemed to be suspicious), 7 known conspirators, and 8 known non-conspirators. We want to identify other members of conspirators and their leaders before arrest them. We first figure out other unknown conspirators and then find the leader by using relationships between conspirators.

For identifying other unknown conspirators, we take all topics into consideration. To one topic, it is ambiguous and uncertain for whether it is a conspiracy or not. However, we can calculate the conspiracy probability of each topic through known conspirators' message traffic based on principle of fuzzy mathematics. Then we get conspirator probability of each member by topics of each one discussed. From method of crime and modus object, we find out the keywords connected with conspiracy. Based on text analysis, we calculate weight of each keyword. The node messages contain more topics connected with conspiracy, the node more probable be conspirator. We put the results of two methods together and compare them. Finally we pick out the unknown conspirators.

For the determination of the leaders, we will determine the accomplice out a separate analysis, first construct a network diagram of these co-conspirators from the figure to identify the most wide coverage or degree of the largest point, the point is the leaders.

2. Assumptions

The topics talked between conspirators are mainly to conspiracy.

The key words we find are all related to conspiracy and conspirator.

The crime form of conspirators is fit in conditions.

Conspirators are all discussed conspiracy in statistical information.

3. Symbols And Significance

x_{i1} The times that conspirators send topic of i

x_{i2} The times that conspirators receive topic of i

t_i The appearance times of topic i

x_i He probability of conspiracy for topic i

$y_{i1,j}$ The times of sending topic i from member j

$y_{i2,j}$ The times of receiving topic i from member j

p_i The probability of conspirators for member i

p_j The weights of key words j

4. To EZ Case

The way of identifying people in the office complex who are the most likely conspirators are based on principle of fuzzy mathematics[1].. We adopt the method of combining qualitative and quantitative. First of all, we analyze simple EZ case which had only 10 people (nodes), 27 links (messages), 5 topics, 1 suspicious/conspiracy topic, 2 known conspirators, and 2 known non-conspirators, as Figure1 shown.

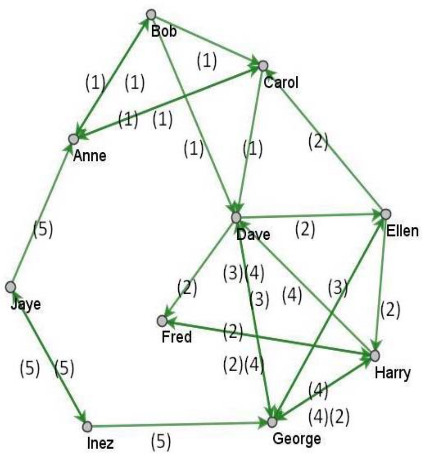


Figure 1 Network of Messages from EZ Case

We number 10 people. The result is shown in Table1

Table 1 Number of 10 Numbers

Name	Anne	Bob	Carol	Dave	Ellen
Number	1	2	3	4	5
Name	Fred	George	Harry	Inez	Jaye
Number	6	7	8	9	10

4.1 Basic Model

As we all known, the more frequent the topic be said by conspirator know， the more probable it is conspiracy. Because Dave and George are known conspirators (NO.4 and NO.7), we can calculate the probability of a topic for whether it is a conspiracy according to the emerged probability of NO.4 and NO.7 in 5 topics. We define the times that conspirators send topic of i as x_{i1} and receive topic as x_{i2} .The number of occurrences for topic i is t_i . Then the probability of conspiracy for topic i we defined is x_i. The relation between x_{i1} , x_{i2} , t_i and x_i is shown as follow:

$$x_i = \frac{x_{i1} + x_{i2}}{t_i} \qquad i = 1,2,3,4,5$$

So, the topic for conspiracy of probability matrix is:

$$X = \left[x_1,\ x_2,\ x_3,\ x_4,\ x_5 \right]$$

We use y_{i1j} expressing the times of sending topic i from member j and y_{i2j} expressing the times of receiving topic i from member j. So the times for every member discussion of each topic y_{ij} is the sum of y_{i1j} and y_{i2j}.

So the matrix for times of discussion is:

$$Y = \begin{bmatrix} y_{11} & \cdots & y_{1j} \\ \vdots & \ddots & \vdots \\ y_{i1} & \cdots & y_{ij} \end{bmatrix} \qquad i = 1,2...,5, j = 1,2,...,10$$

According to the matrix Y above, we can gain the total times of each member sending and receiving topic:

$$Z = \left[\sum_{i=1}^{5} y_{i1}, \sum_{i=2}^{5} y_{i2}, \cdots\cdots, \sum_{i=5}^{5} y_{ij} \right]$$

If we let p_j represent as the probability of conspirators for member j then we can get: $p_j = \frac{\sum_{i=1}^{5} y_{ij} \times x_i}{\sum_{i=1}^{5} y_{ij}}$

$$\mathrm{P} = \left[p_1, p_2, p_3, \cdots\cdots, p_j \right]$$

Normalization processing:

$$P' = \frac{p_j - \min(P)}{\max(P) - \min(P)}$$

The result is shown in Table2

Table 2 Probability of 10 Members on Basic Model

Number	1	2	3	4	5
Probability	0.205	0.220	0.233	0.750	0.703
Number	6	7	8	9	10
Probability	0.416	1	0.455	0	0

From Table2, we can get a conclusion that member 4, 5, 7 are conspirators, member 1, 2, 3, 9, 10 are non-conspirators and member 6, 8 are unsure. However, it is not fit the fact that member 2, 4, 5, 7, 9 are conspirators, member 1, 3, 10 are non- conspirators and member 6, 8 are unsure.

4.2 Improved Method

The weakness which we calculated the probability of conspiracy in topic before is that we considered sending messages together with receiving messages. In fact, the effect of two aspects is different. So we should consider them separately. We assume α as the degree of effect on conspiracy in topic when conspirators send messages and β as the degree of effect on conspiracy in topic when conspirators receive messages. Obviously, we can gain $\alpha+\beta=1$ .The conspiracy probability of topic i changes into:

$$x_i = \frac{\alpha x_{i1} + \beta x_{i2}}{t_i} \qquad i = 1,2,3,4,5$$

To estimate parameter α, β we regulate the value of them by calculating P' and comparing P' with truth. Try many times. We conclude that it is appropriate when $\alpha = 0.2$ and $\beta = 0.8$. So we conclude:

$$x_i = \frac{0.2x_{i1} + 0.8x_{i2}}{t_i}$$

The result is shown in Table3

Table 3 Probability of 10 Members on Improved Model

Number	1	2	3	4	5
Probability	0.101	0.112	0.090	0.771	0.538
Number	6	7	8	9	10
Probability	0	1	0.424	0.056	0.056

The actual results of Bob, Dave, Ellen, George, Inez is an accomplice, Anne, Carol, Jaye is not an accomplice, comparison is not particularly close to the results., This is because the data is not enough.

5. To The Current Case

The current case has 83 nodes, 400 links (some involving more than 1 topic), over 21,000 words of message traffic,15 topics (3 have been deemed to be suspicious), 7 known conspirators, and 8 known non-conspirators

5.1 Using Improved Model

According to our improved model, we calculate the probability of each member being conspirator. The outcome is sorted in ascending order and shown in Table4

Table 4 Probability of 83 Members on Improved Model

N	P	N	P	N	P	N	P	N	P
53	0.000	41	0.33	50	0.409	23	0.481	43	0.577
57	0.000	25	0.331	62	0.412	12	0.485	37	0.579
59	0.000	82	0.331	64	0.412	39	0.487	13	0.621
61	0.093	24	0.341	42	0.412	79	0.49	16	0.646
77	0.228	55	0.346	5	0.419	20	0.491	22	0.647
80	0.228	36	0.355	15	0.422	3	0.503	49	0.655
68	0.258	10	0.358	0	0.425	29	0.517	9	0.679
58	0.259	1	0.369	35	0.425	8	0.522	7	0.685
63	0.259	6	0.371	32	0.448	31	0.522	47	0.697
74	0.289	28	0.395	2	0.453	65	0.524	21	0.706
17	0.293	66	0.398	71	0.455	19	0.525	67	0.779
70	0.301	73	0.398	34	0.461	40	0.533	54	0.823
14	0.308	11	0.399	52	0.461	78	0.537	81	0.844
30	0.317	48	0.399	60	0.468	18	0.557	51	0.980
76	0.318	4	0.401	45	0.474	33	0.561	56	1.000
26	0.325	38	0.401	75	0.477	44	0.568		
69	0.326	72	0.407	46	0.478	27	0.576		

(N: on behalf of serial number. P: on behalf of conspiracy probability. The shaded means known conspirators)

From Table4 we know that the conspiracy probability of known conspirators is bigger than others. That is consistent with fact.

5.2 The Semantic Network Analysis Model

In the previous method, we stared from known conspirators and figured out the probability of each topic being conspiracy. And then we in turn calculated the probability of who may be conspirator. The next we reconsider from aspects of words on message traffic. If someone's messages are mostly connected to the conspiracy, he is more likely to be a conspirator. This idea comes from the method of text analysis.

5.2.1 Background

With the development of computer network technology, the exchange between people becomes more and more convenient. The semantic analysis and text analysis become increasingly important and difficult. Recently text analysis has focused on text representation model selection and the selection of feature selection algorithm. Our model is based on the semantic network analysis.

5.2.2 Analysis Model

By known conditions, a conspiracy is taking place to embezzle funds from the company and use internet fraud to steal funds from credit cards of people who do business with the company. From the details about the topics in file Topics.xls, we find key words from suspicious message topics. We first find three key words: Spanish, Paige and Compute. If someone has more keywords in his message, he has more conspiracy probability.

We use formula of TFIDF (term frequency–inverse document frequency) to figure out weights of keywords that is based on principle of semantic analysis for solving weight [2]-[3]. The formula:

$$W(i,j) = LW(i,j) * GWT(i)$$

$LW(i,j) = tf_{ij}$ as local weight, $GWT(i) = df_i / N$ as global weight, tf_{ij} as frequency of key word i appearing in message j, df_i as amounts of message appearing key word i and N as total number of message.

Because the more frequent key words appearing the more probable being conspirator, we gain the conspirator probability p_k of member k:

$$p_k = \frac{\sum W(i,j) * t_{kj}}{\sum t_{kj}}$$

t_{kj} as the sending or receiving times for topic j. Here we have some innovation.

The result is not ideal through this way, for example Jean and Yao who are known conspirators but out of conspirators list in our result. Analysis again to our model, we find that the reason is our key words is too less. So we add up another key word "finance" which is from Topic 1 (one of condition changes). Try again as before, the result is shown in Table5

Table 5 Probability of 83 Members on Semantic Network Analysis Model

N	P	N	P	N	P	N	P	N	P
53	0.000	26	0.338	5	0.428	45	0.499	27	0.596
57	0.000	41	0.339	50	0.431	3	0.510	44	0.597
59	0.000	62	0.340	15	0.441	52	0.516	13	0.647
61	0.093	64	0.340	42	0.443	20	0.52	16	0.651
77	0.228	82	0.340	71	0.446	39	0.522	49	0.66
80	0.228	69	0.341	0	0.451	78	0.537	22	0.671
68	0.247	24	0.356	35	0.458	8	0.543	21	0.685
58	0.259	10	0.360	32	0.465	19	0.545	7	0.696
63	0.259	36	0.378	60	0.483	65	0.547	9	0.702
74	0.289	1	0.392	34	0.483	12	0.548	47	0.715
17	0.293	28	0.395	46	0.487	29	0.549	67	0.805
14	0.303	6	0.397	79	0.49	31	0.552	54	0.808
55	0.310	72	0.407	23	0.496	40	0.566	81	0.844
70	0.314	38	0.414	66	0.497	18	0.576	51	0.980
76	0.318	11	0.418	73	0.497	43	0.577	56	1.000
30	0.321	48	0.425	2	0.498	33	0.591		
25	0.324	4	0.428	75	0.499	37	0.593		

(N: on behalf of serial number. P: on behalf of conspiracy probability. The shaded means known conspirators)

If $P_K \geq \min\{P_7, P_{18}, P_{21}, P_{37}, P_{43}, P_{49}, P_{54}, P_{67}\}$,then k is an accomplice.

From Table5, we conclude that conspirators are Elsie, Malcolm, Marion, Jerome, Jean, Alex, Eric, Marcia, Elsie, Paul, Christina, Harvey, Dayi, Ulf, Cha, Yao and Seeni. For Jean, Alex, Elsie, Paul, Ulf, Yao and Harvey are known conspirators.

From the Names.xls, we find some members have same name, for example NO.16 and NO.34 have the same name

"Jerome", NO.4 and NO.32 are all named "Gretchen". To solving the problem, we express all names in other way such as Jerome16, Jerome34, Delores10, Gretchen4 and Gretchen32. Their probability respectively: 0.646, 0.461, 0.358, 0.401, 0.448. Since Jerome16 is highest so we can say Jerome16 is most likely to be conspirator.

5.3 Explore The Model on Conditions Change

The conditions change: Topic 1 is also connected to the conspiracy and that Chris is one of the conspirators. The same method we used as former, the result is shown in Table6

Table 6 Probability of 83 Members on Conditions Change

N	P	N	P	N	P	N	P	N	P
53	0.000	82	0.286	64	0.369	34	0.436	16	0.549
57	0.000	25	0.305	15	0.374	0	0.437	43	0.557
59	0.000	14	0.308	10	0.374	50	0.439	13	0.561
61	0.053	69	0.316	24	0.382	12	0.45	49	0.561
55	0.210	76	0.316	71	0.386	48	0.454	27	0.579
68	0.210	26	0.325	75	0.395	52	0.456	7	0.605
58	0.211	42	0.329	38	0.399	40	0.489	9	0.605
63	0.211	70	0.333	19	0.400	60	0.491	22	0.618
17	0.228	45	0.337	2	0.405	3	0.508	21	0.632
36	0.253	28	0.342	20	0.408	33	0.509	47	0.662
30	0.257	72	0.342	6	0.412	31	0.511	54	0.716
41	0.263	35	0.353	46	0.412	44	0.518	67	0.750
66	0.263	4	0.369	32	0.415	65	0.518	81	0.768
73	0.263	5	0.369	39	0.421	78	0.526	56	0.790
74	0.263	11	0.369	79	0.421	8	0.542	51	1.000
77	0.263	23	0.369	1	0.429	18	0.542		
80	0.263	62	0.369	29	0.434	37	0.547		

(N: on behalf of serial number. P: on behalf of conspiracy probability. The shaded means known conspirators)

Compared to the probability in different conditions, we can see they have little difference. The result is shown in Figure2.

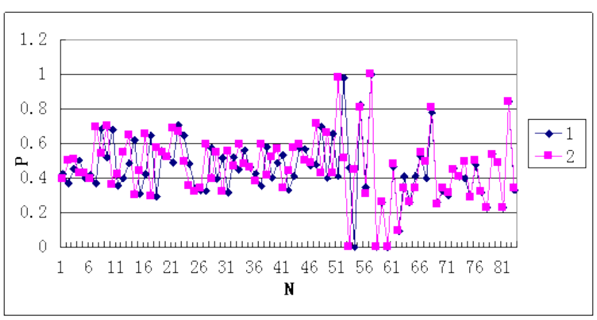


Figure 2 The Probability in Different Conditions

6. The Network Sociology Analysis

6.1 Background

1930, W. L. Warner pointed out that the social structure of a modern community type constitute by many sub-groups, such as family, church and classes.[4]
1972, Bruce Kapferer successfully predicted a strike. That greatly improved the level of theory and practice of network sociology. [5]
Network sociology may be subordinated to the future independent discipline network science (Weizhi Deng 2001) [6]
We propose a network sociology model to nominate the conspiracy leaders that is based on network analysis. The model is run in UCINET software which is one of most popular simple software of social network analysis at the present time.
Known by the common sense, it is very useful to combat the conspiracy leaders for fighting against criminal gangs. The leading figure is the hub of the network for

information exchange based on the social network model. If someone's degree centrality and betweenness centrality is rank in front of sequence we think he is the conspiracy leader. So we calculate the parameter of degree centrality and betweenness centrality to find the conspiracy leaders.

6.2 The Relationships Matrix and Network Diagram

We use 1 and 0 to describe the two whether linked or not and build the relationships matrix for conspirators. Then we draw the relationship network diagram by using UCINET software. The diagram is shown in Figure3

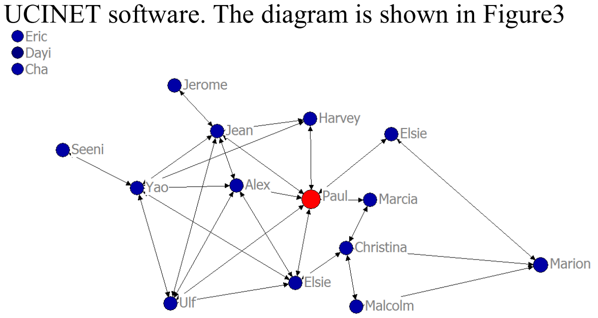


Figure 3 Network Diagram of Conspirators

6.3 The Degree Centrality and The Betweenness Centrality

For a network node, degree is the most basic connectivity metric parameters. Degree is expressed by the node and other nodes connection number d and divided into in-degree and out-degree. 1974, Nieminen put forward the calculation formula of degree centrality:[7]

$$C_D = \sum_{i=1}^{n} \alpha(p_i, p_k)$$

Table 7 The Degree Centrality of Conspirators

Number	Name	Degree	Number	Name	Degree
10	Paul	7.000	9	Elsie	2.000
5	Jean	6.000	8	Marcia	2.000
16	Yao	6.000	2	Malcolm	2.000
1	Elsie	5.000	17	Seeni	1.000
6	Alex	5.000	4	Jerome	1.000
14	Ulf	5.000	7	Eric	0.000
11	Christina	4.000	15	Cha	0.000
3	Marion	3.000	13	Dayi	0.000
12	Harvey	3.000			

We can see that Paul, Jean and Yao are the top three. Paul is the most probable conspirator leader whose degree centrality is 7.00.

Table 8 The Betweenness Centrality of Conspirators

Number	Name	Betweenness	Number	Name	Betweenness
10	Paul	25.050	8	Marcia	2.150
1	Elsie	17.400	12	Harvey	1.133
16	Yao	14.983	7	Eric	0.000
11	Christina	14.917	13	Dayi	0.000
5	Jean	13.800	2	Malcolm	0.000
9	Elsie	5.917	15	Cha	0.000
3	Marion	2.917	4	Jerome	0.000
6	Alex	2.367	17	Seeni	0.000
14	Ulf	2.367			

From the table above, Paul is also in the top and his betweenness centrality is 20.05.
Because Paul's degree centrality and the betweenness centrality is the highest from others.Paul is the most probable to be conspirator leader. In summary, Paul is the conspirator leader.

7. Model Evaluation

7.1 Model Promotion

The former models all took one aspect of factor into consideration. Basic model and improved model only consider "Conspirators". The semantic network analysis model and the network sociology model only take "Key words" into account. To improving our model, we consider four factors "Conspirators", "Key words", "Non-conspirators", "Suspicious topics". We make a comprehensive evaluation with four factors. The improvement ideas graph is shown in Figure4

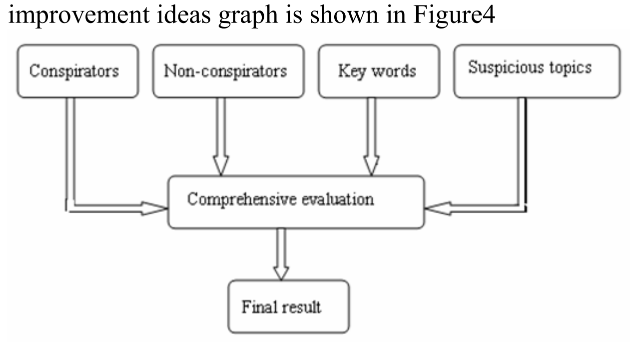


Figure 4 The Improvement Ideas Graph

The model improved could apply to the assessment of product quality, evaluation of the quality of hotel services, and also could apply to cluster analysis.

7.2 Strength

Basic model and improved model: we adopt principle of fuzzy mathematics. We start from known conspirators. We deal with the data and calculate the weight of topic contain conspirators node's boundary x_i,we regard this probability as conspiracy topic. Then we calculate in turn the conspirator probability of nodes. So, we quantify each node and it is easy to sorting, comparison and screening conspirators.

The semantic network analysis mode: we take conspiracy topic into account and pick out key word connected with conspiracy. We calculate weight with semantic analysis. Taking consideration from key words is close to our purpose and avoiding leaving out some conspirator when people discuss too much insignificant topic.

7.3 Weakness

Basic model and improved model: It is a bit one-sided to estimate the probability of conspiracy by frequency of conspiracy occurrence. If we comprehensive evaluate three factors "Conspirators", "Non-conspirators", and "Suspicious topics", the result will be better.

The semantic network analysis mode: The keywords selection and number determined directly determine the accuracy of the results. If the first step of selection keywords are wrong it could result in incalculable error on the overall.

Reference

[1] Guangwu Meng. The Basic Theory and Its Application of Fuzzy Mathematics(1)——Overview of The Emergence and Development of Fuzzy Math. Liaocheng Teachers University (Natural Science Edition), 1998, (6)
[2] Yunfeng Liu, Huan Qi, Xiang'en Hu and Zhiqiang Cai. Weight Aalculated By Latent Semantic Analysis Improvements. Journal of Chinese Information Processing, Vol. 19 No.6: 64-69
[3] Lei Chen, Bidan Dong and Yanping Zhao. Hidden Within The Social Network Analysis in The Semantic - based Enterprise Relationship Detection. Computer and Digital Engineering, 2009, (9): 58-63
[4] John Scott. Social Network Analysis [M]. Chong Qing: Chongqing University Press, 2007
[5] Martin kilduff, Wenpin Tsai. Social Networks and Organizations [M]. Sage Publications of London, 2003
[6] Weizhi Deng, Jianwei Fan and Weisheng Shi. On The Establishment of The Network of Community Schools [J]. Jianghai Tribune Phase IV, 2001
[7] Freeman, L.C. Centrality in Social Networks I: Conceptual Clarification [J]. Social Networks, 1979, (1):215-239.
[8] Zeng Xianzhao. Network Science :Volume II. Bei Jing: Military Science Press. 112-162(2008).

First Author: Zhi-hong Ma (1975-), male, born in Ningxia, Associate Professor of the Tianjin agricultural University, master's degree, mainly engaged in the teaching of mathematics and applied mathematics research.

Second Author Xun-song He is a student of Tianjin Agricultural University .

Third Author Hao-xuan Ding is a student of Tianjin Agricultural University .

IJCSI CALL FOR PAPERS JANUARY 2014 ISSUE

Volume 11, Issue 1

The topics suggested by this issue can be discussed in term of concepts, surveys, state of the art, research, standards, implementations, running experiments, applications, and industrial case studies. Authors are invited to submit complete unpublished papers, which are not under review in any other conference or journal in the following, but not limited to, topic areas. See authors guide for manuscript preparation and submission guidelines.

Accepted papers will be published online and indexed by Google Scholar, Cornell's University Library, DBLP, ScientificCommons, CiteSeerX, Bielefeld Academic Search Engine (BASE), SCIRUS, EBSCO, ProQuest and more.

Deadline: 15th January 2014
Online Publication: 31st January 2014

- Evolutionary computation
- Industrial systems
- Evolutionary computation
- Autonomic and autonomous systems
- Bio-technologies
- Knowledge data systems
- Mobile and distance education
- Intelligent techniques, logics, and systems
- Knowledge processing
- Information technologies
- Internet and web technologies
- Digital information processing
- Cognitive science and knowledge agent-based systems
- Mobility and multimedia systems
- Systems performance
- Networking and telecommunications
- Software development and deployment
- Knowledge virtualization
- Systems and networks on the chip
- Context-aware systems
- Networking technologies
- Security in network, systems, and applications
- Knowledge for global defense
- Information Systems [IS]
- IPv6 Today - Technology and deployment
- Modeling
- Optimization
- Complexity
- Natural Language Processing
- Speech Synthesis
- Data Mining

For more topics, please see http://www.ijcsi.org/call-for-papers.php

All submitted papers will be judged based on their quality by the technical committee and reviewers. Papers that describe on-going research and experimentation are encouraged.
All paper submissions will be handled electronically and detailed instructions on submission procedure are available on IJCSI website (www.IJCSI.org).

For more information, please visit the journal website (www.IJCSI.org)

IJCSI

The International Journal of Computer Science Issues (IJCSI) is a well-established and notable venue for publishing high quality research papers as recognized by various universities and international professional bodies. IJCSI is a refereed open access international journal for publishing scientific papers in all areas of computer science research. The purpose of establishing IJCSI is to provide assistance in the development of science, fast operative publication and storage of materials and results of scientific researches and representation of the scientific conception of the society.

It also provides a venue for researchers, students and professionals to submit ongoing research and developments in these areas. Authors are encouraged to contribute to the journal by submitting articles that illustrate new research results, projects, surveying works and industrial experiences that describe significant advances in field of computer science.

Indexing of IJCSI

1. Google Scholar
2. Bielefeld Academic Search Engine (BASE)
3. CiteSeerX
4. SCIRUS
5. Docstoc
6. Scribd
7. Cornell's University Library
8. SciRate
9. ScientificCommons
10. DBLP
11. EBSCO
12. ProQuest