IJCSIS Vol. 14 No. 3, March 2016 Part II
ISSN 1947-5500

International Journal of Computer Science & Information Security

Indexed and technically co-sponsored by :

IJCSIS

ISSN (online): 1947-5500

Please consider to contribute to and/or forward to the appropriate groups the following opportunity to submit and publish original scientific results.

CALL FOR PAPERS
International Journal of Computer Science and Information Security (IJCSIS)
January-December 2016 Issues

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 Authors are invited to submit complete unpublished papers, which are not under review in any other conference or journal in the following, but not limited to, topic areas.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.

Indexed by Google Scholar, DBLP, CiteSeerX, Directory for Open Access Journal (DOAJ), Bielefeld Academic Search Engine (BASE), SCIRUS, Scopus Database, Cornell University Library, ScientificCommons, ProQuest, EBSCO and more.

Deadline: **see web site**
Notification: see web site
Revision: see web site
Publication: see web site

Context-aware systems
Networking technologies
Security in network, systems, and applications
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
Knowledge for global defense
Information Systems [IS]
IPv6 Today - Technology and deployment
Modeling
Software Engineering
Optimization
Complexity
Natural Language Processing
Speech Synthesis
Data Mining

For more topics, please see web site https://sites.google.com/site/ijcsis/

For more information, please visit the journal website (https://sites.google.com/site/ijcsis/)

Editorial
Message from Editorial Board

*It is our great pleasure to present the **March 2016 issue** (Volume 14 Number 3) of the **International Journal of Computer Science and Information Security** (IJCSIS). High quality survey and review articles are proposed from experts in the field, promoting insight and understanding of the state of the art, and trends in computer science and technology. The contents include original research and innovative applications from all parts of the world. According to Google Scholar, up to now papers published in IJCSIS have been cited over 5668 times and the number is quickly increasing. This statistics shows that IJCSIS has established the first step to be an international and prestigious journal in the field of Computer Science and Information Security. The main objective is to disseminate new knowledge and latest research for the benefit of all, ranging from academia and professional communities to industry professionals. It especially provides a platform for high-caliber researchers, practitioners and PhD/Doctoral graduates to publish completed work and latest development in active research areas. IJCSIS is indexed in major academic/scientific databases and repositories: Google Scholar, CiteSeerX, Cornell's University Library, Ei Compendex, ISI Scopus, DBLP, DOAJ, ProQuest, Thomson Reuters, ArXiv, ResearchGate, Academia.edu and EBSCO among others.*

On behalf of IJCSIS community and the sponsors, we congratulate the authors and thank the reviewers for their dedicated services to review and recommend high quality papers for publication. In particular, we would like to thank the international academia and researchers for continued support by citing papers published in IJCSIS. Without their sustained and unselfish commitments, IJCSIS would not have achieved its current premier status.

*"We support researchers to succeed by providing high visibility & impact value, prestige and excellence in research publication." For further questions or other suggestions please do not hesitate to contact us at **ijcsiseditor@gmail.com**.*

A complete list of journals can be found at:
http://sites.google.com/site/ijcsis/

IJCSIS Vol. 14, No. 3, March 2016 Edition

ISSN 1947-5500 © IJCSIS, USA.

Journal Indexed by (among others):

Bibliographic Information
ISSN: 1947-5500
Monthly publication (Regular Special Issues)
Commenced Publication since May 2009

Editorial / Paper Submissions:
IJCSIS Managing Editor
(ijcsiseditor@gmail.com)
Pennsylvania, USA
Tel: +1 412 390 5159

TABLE OF CONTENTS

Sabrina Nefti, Dept. of Computer Science, University Batna 2, Algeria
Mammar Sedrati, Dept. of Computer Science, University Batna 2, Algeria

Abstract — Advances in multimedia and ad-hoc networking have urged a wealth of research in multimedia delivery over ad-hoc networks. This comes as no surprise, as those networks are versatile and beneficial to a plethora of applications where the use of fully wired network has proved intricate if not impossible, such as prompt formation of networks during conferences, disaster relief in case of flood and earthquake, and also in war activities. It this paper, we aim to investigate the combined impact of network sparsity and network node density on the Peak Signal Noise to Ratio (PSNR) and jitter performance of proactive and reactive routing protocols in ad-hoc networks. We also shed light onto the combined effect of mobility and sparsity on the performance of these protocols. We validate our results through the use of an integrated Simulator-Evaluator environment consisting of the Network Simulator NS2, and the Video Evaluation Framework Evalvid.

Keywords- PSNR, MANET, Sparsity, Density, Routing protocols, Video Streaming, NS2, Evalvid

Mahmoud Al-Ayyoub, Ala'a Oqaily and Mohammad I. Jarrah
Jordan University of Science and Technology Irbid, Jordan
Huda Karajeh, The University of Jordan Amman, Jordan

Abstract — Computer-aided diagnosis (CAD) systems have gained a lot of popularity in the past few decades due to their effectiveness and usefulness. A large number of such systems are proposed for a wide variety of abnormalities including those related to coronary artery disease. In this work, a CAD system is proposed for such a purpose. Specifically, the proposed system determines the location of thrombosis in x-ray coronary angiograms. The problem at hand is a challenging one as indicated by some researchers. In fact, no prior work has attempted to address this problem to the best of our knowledge. The proposed system consists of four stages: image preprocessing (which involves noise removal), vessel enhancement, segmentation (which is followed by morphological operations) and localization of thrombosis (which involves skeletonization and pruning before localization). The proposed system is tested on a rather small dataset and the results are encouraging with a 90% accuracy.

Keywords — Heterogeneous wireless networks, Vertical handoff, Markov model, Artificial intelligence, Mobility management.

Carlos André Batista de Carvalho (#∗) , Rossana Maria de Castro Andrade (∗), Márcio E. F. Maia (∗) , Davi Medeiros Albuquerque (∗) , Edgar Tarton Oliveira Pedrosa (∗)
 # Computer Science Department, Federal University of Piaui, Brazil
** Group of Computer Networks, Software Engineering, and Systems, Federal University of Ceara, Brazil*

Abstract — Mobile devices became a natural target of security threats due their vast popularization. That problem is even more severe when considering Android platform, the market leader operating system, built to be open and extensible. Although Android provides security countermeasures to handle mobile threats, these defense measures are not sufficient and attacks can be performed in this platform, exploiting existing vulnerabilities. Then, this paper focuses on improving the security of the Android ecosystem with a contribution that is two-fold, as follows: i) a process to analyze and mitigate Android vulnerabilities, scrutinizing existing security breaches found in the literature and proposing mitigation actions to fix them; and ii) an experience report that describes four vulnerabilities and their corrections, being one of them a new detected and mitigated vulnerability.

4. Paper 29021655: Performance Analysis of Proposed Network Architecture: OpenFlow vs. Traditional Network (pp. 30-39)

Idris Z. Bholebawa (#), Rakesh Kumar Jha (*), Upena D. Dalal (#)
(#) Department of Electronics and Communication Engineering, S. V. National Institute of Technology, Surat, Gujarat, India.
(*) School of Electronics and Communication Engineering, Shri Mata Vaishno Devi University, Katra, J&K

Abstract – The Internet has been grown up rapidly and supports variety of applications on basis of user demands. Due to emerging technological trends in networking, more users are becoming part of a digital society, this will ultimately increases their demands in diverse ways. Moreover, traditional IP-based networks are complex and somehow difficult to manage because of vertical integration problem of network core devices. Many research projects are under deployment in this particular area by network engineers to overcome difficulties of traditional network architecture and to fulfill user requirements efficiently. A recent and most popular network architecture proposed is Software-Defined Networks (SDN). A purpose of SDN is to control data flows centrally by decoupling control plane and data plane from network core devices. This will eliminate the difficulty of vertical integration in traditional networks and makes the network programmable. A most successful deployment of SDN is OpenFlow-enabled networks.
In this paper, a comparative performance analysis between traditional network and OpenFlow-enabled network is done. A performance analysis for basic and proposed network topologies is done by comparing round-trip propagation delay between end nodes and maximum obtained throughput between nodes in traditional and OpenFlow-enabled network environment. A small campus network have been proposed and performance comparison between traditional network and OpenFlow-enabled network is done in later part of this paper. An OpenFlow-enabled campus network is proposed by interfacing virtual node of virtually created OpenFlow network with real nodes available in campus network. An implementation of all the OpenFlow-enabled network topologies and a proposed OpenFlow-enabled campus network is done using open source network simulator and emulator called Mininet. All the traditional network topologies are designed and analyzed using NS2 - network simulator.

Keywords – SDN, OpenFlow, Mininet, Network Topologies, Interfacing Network.

5. Paper 29021622: Reverse Program Analyzed with UML Starting from Object Oriented Relationships (pp. 40-45)

Hamed J. Al-Fawareh, Software Engineering Department, Zarka University, Jordan

Abstract - In this paper, we provide a reverse-tool for object oriented programs. The tool focuses on the technical side of maintaining object-oriented program and the description of associations graph for representing meaningful diagram between components of object-oriented programs. In software maintenance perspective reverse engineering process extracts information to provide visibility of the object oriented components and relations in the software that are essential for maintainers.

Keywords: Software Maintenance, Reverse Engineering.

6. Paper 29021628: Lifetime Optimization in Wireless Sensor Networks Using FDstar-Lite Routing Algorithm (pp. 46-55)

Imad S. Alshawi, College of Computer Science and Information Technology, Basra University, Basra, Iraq
Ismaiel O. Alalewi, College of Science, Basra University, Basra, Iraq

Abstract — Commonly in Wireless Sensor Networks (WSNs), the biggest challenge is to make sensor nodes that are energized by low-cost batteries with limited power run for longest possible time. Thus, energy saving is indispensible concept in WSNs. The method of data routing has a pivotal role in conserving the available energy since remarkable amount of energy is consumed by wireless data transmission. Therefore, energy efficient routing protocols can save battery power and give the network longer lifetime. Using complex protocols to plan data routing efficiently can reduce energy consumption but can produce processing delay. This paper proposes a new routing method called FDstar-Lite which combines Dstar-Lite algorithm with Fuzzy Logic. It is used to find the optimal path from the source node to the destination (sink) and reuse that path in such a way that keeps energy consumption fairly distributed over the nodes of a WSN while reducing the delay of finding the routing path from scratch each time. Interestingly, FDstar-Lite was observed to be more efficient in terms of reducing energy consumption and decreasing end-to-end delay when compared with A-star algorithm, Fuzzy Logic, Dstar-Lite algorithm and Fuzzy A-star. The results also show that, the network lifetime achieved by FDstar-Lite could be increased by nearly 35%, 31%, 13% and 11% more than that obtained by A-star algorithm, Fuzzy Logic, Dstar-Lite algorithm and Fuzzy A-star respectively.

Keywords— Dstar-Lite algorithm, fuzzy logic, network lifetime, routing, wireless sensor network.

7. Paper 29021637: An Algorithm for Signature Recognition Based on Image Processing and Neural Networks (pp. 56-60)

Ramin Dehgani, Ali Habiboghli
Department of computer science and engineering, Islamic Azad University, Khoy, Iran

Abstract — Characteristics related to people signature has been extracted in this paper. Extracted Specialty vector under neural network has been used for education. After teaching network, signatures have been evaluated by educated network to recognize real signature from unreal one. Comparing the results shows that the efficiency of this method is better than the other methods.

Index Terms— signature recognition, neural networks, image processing.

8. Paper 29021640: A Report on Using GIS in Establishing Electronic Government in Iraq (pp. 61-64)

Ahmed M.JAMEL, Department of Computer Engineering, Erciyes University, Kayseri, Turkey
Dr. Tolga PUSATLI, Department of Mathematics and Computer Science, Cankaya University, Ankara, Turkey

Abstract — Electronic government initiatives and public participation in them are among the indicators of today's development criteria for countries. After the consequent of two wars, Iraq's current position in, for example, the UN's e-government ranking is quite low and did not improve in recent years. In the preparation of this work, we are motivated by the fact that handling geographic data of the public facilities and resources are needed in most of the e-government projects. Geographical information systems (GIS) provide the most common tools, not only to manage spatial data, but also to integrate with non-spatial attributes of the features. This paper proposes that establishing a working GIS in the health sector of Iraq would improve e-government applications. As a case study, investigating hospital locations in Erbil has been chosen. It is concluded that not much is needed to start building base works for GIS supported e-government initiatives.

Keywords - Electronic government, Iraq, Erbil, GIS, Health Sector.

9. Paper 29021642: Satellite Image Classification by Using Distance Metric (pp. 65-68)

Dr. Salem Saleh Ahmed Alamri, Dr. Ali Salem Ali Bin-Sama
Department of Engineering Geology, Oil & Minerals Faculty, Aden University, Aden, Yemen
Dr. Abdulaziz Saleh Yeslam Bin–Habtoor
Department of Electronic and Communication Engineering, Faculty of Engineering & Petrolem, Hadramote University, Mokula , Yemen

Abstract — This paper attempts to undertake the study satellite image classification by using six distance metric as Bray Curtis Distance Method, Canberra Distance Method, Euclidean Distance Method, Manhattan Distance Method, Square Chi Distance Method, Squared Chord Distance Method and they are compared with one another, So as to choose the best method for satellite image classification.

Keyword: Satellite Image, Classification, Texture Image, Distance Metric,

10. Paper 29021650: Cybercrime and its Impact on E-government Services and the Private Sector in The Middle East (pp. 69-73)

Sulaiman Al Amro, Computer Science (CS) Department, Qassim University, Buraydah, Qassim, 51452, KSA

Abstract — This paper will discuss the issue of cybercrime and its impact on both e-government services and the private sector in the Middle East. The population of the Middle East has now become increasingly connected, with ever greater use of technology. However, the issue of piracy has continued to escalate, without any signs of abating. Acts of piracy have been established as the most rapidly growing (and efficient) sector within the Middle East, taking advantage of attacks on the infrastructure of information technology. The production of malicious software and new methods of breaching security has enabled both amateur and professional hackers and spammers, etc., to target the Internet in new and innovative ways, which are, in many respects, similar to legitimate businesses in the region.

Keywords - cybercrimes; government sector; private sectors; Middle East; computer security

11. Paper 29021657: Performance Comparison between Forward and Backward Chaining Rule Based Expert System Approaches Over Global Stock Exchanges (pp. 74-81)

Sachin Kamley, Deptt. of Computer Application's S.A.T.I., Vidisha, India
Shailesh Jaloree, Deptt. of Appl. Math's and CS S.A.T.I., Vidisha, India
R.S. Thakur, Deptt. of Computer Application's M.A.N.I.T., Bhopal, India

Abstract — For the last couple of decade's stock market has been considered as a most noticeable research area everywhere throughout the world because of the quickly developing of the economy. Throughout the years, a large portion of the researchers and business analysts have been contributed around there. Extraordinarily, Artificial Intelligence (AI) is the principle overwhelming area of this field. In AI, an expert system is one of the understood and prevalent techniques that copy the human abilities in order to take care of particular issues. In this research study, forward and backward chaining two primary expert system inference methodologies is proposed to stock market issue and Common LISP 3.0 based editors are used for designing an expert system shell. Furthermore, expert systems are tested on four noteworthy global stock exchanges, for example, India, China, Japan and United States (US). In addition, different financial components, for example, Gross Domestic Product (GDP), Unemployment Rate, Inflation Rate and Interest Rate are also considered to build the expert knowledge base system. Finally, experimental results demonstrate that the backward chaining approach has preferable execution performance over forward chaining approach.

Keywords— Stock Market; Artificial Intelligence; Expert System; Macroeconomic Factors; Forward Chaining; Backward Chaining; Common LISP 3.0.

12. Paper 29021658: Analysis of Impact of Varying CBR Traffic with OLSR & ZRP (pp. 82-85)

Rakhi Purohit, Department of Computer Science & Engineering, Suresh Gyan Vihar University, Jaipur, Rajasthan, India
Bright Keswani, Associate Professor & Head Department of Computer Application, Suresh Gyan Vihar University, Jaipur, Rajasthan, India

Abstract — Mobile ad hoc network is the way to interconnect various independent nodes. This network is decentralize and not follows any fixed infrastructure. All the routing functionality are controlled by all the nodes. Here nodes can be volatile in nature so they can change place in network and effect network architecture. Routing in mobile ad hoc network is very much dependent on its protocols which can be proactive and reactive as well as with both features. This work consist of analysis of protocols have analyzed in different scenarios with varying data traffic in the network. Here OLSR protocol has taken as proactive and ZRP as Hybrid protocol. Some of the calculation metrics have evaluated for this analysis. This analysis has performed on well-known network simulator NS2.

Index Terms:- Mobile ad hoc network, Routing, OLSR, Simulation, and NS2.

13. Paper 29021662: Current Moroccan Trends in Social Networks (pp. 86-98)

Abdeljalil EL ABDOULI, Abdelmajid CHAFFAI, Larbi HASSOUNI, Houda ANOUN, Khalid RIFI,
RITM Laboratory, CED Engineering Sciences, Ecole Superieure de Technologie, Hassan II University of Casablanca, Morocco

Abstract — The rapid development of social networks during the past decade has lead to the emergence of new forms of communication and new platforms like Twitter and Facebook. These are the two most popular social networks in Morocco. Therefore, analyzing these platforms can help in the interpretation of Moroccan society current trends. However, this will come with few challenges. First, Moroccans use multiple languages and dialects for their daily communication, such as Standard Arabic, Moroccan Arabic called “Darija”, Moroccan Amazigh dialect called “Tamazight”, French, and English. Second, Moroccans use reduced syntactic structures, and unorthodox lexical forms, with many abbreviations, URLs, #hashtags, spelling mistakes. In this paper, we propose a detection engine of Moroccan social trends, which can extract the data automatically, store it in a distributed system which is the Framework Hadoop using the HDFS storage model. Then we process this data, and analyze it by writing a distributed program with Pig UDF using Python language, based on Natural Language Processing (NLP) as linguistic technique, and by applying the Latent Dirichlet Allocation (LDA) for topic modeling. Finally, our results are visualized using pyLDAvis, WordCloud, and exploratory data analysis is done using hierarchical clustering and other analysis methods.

Keywords: distributed system; Framework Hadoop; Pig UDF; Natural Language Processing; Latent Dirichlet Allocation; topic modeling; pyLDAvis; wordcloud; exploratory data analysis; hierarchical clustering.

14. Paper 29021665: Design Pattern for Multilingual Web System Development (pp. 99-105)

Dr. Habes Alkhraisat, Al Balqa Applied University, Jordan

Abstract — Recently- Multilingual WEB Database system have brought into sharp focus the need for systems to store and manipulate text data efficiently in a suite of natural languages. While some means of storing and querying multilingual data are provided by all current database systems. In this paper, we present an approach for efficient development multilingual web database system with the use of object oriented design principle benefits. We propose functional, efficient, dynamic and flexible object oriented design pattern and database system architecture for making the performance of the database system to be language independent. Results from our initial implementation of the proposed methodology are encouraging indicating the value of proposed approach.

Index Terms— Database System, Design Pattern, Inheritance, Object Oriented, Structured Query Language.

15. Paper 29021669: A Model for Deriving Matching Threshold in Fingerprint-based Identity Verification System (pp. 106-114)

Omolade Ariyo. O., Fatai Olawale. W.
Department of Computer Science, University of Ilorin, Ilorin, Nigeria

Abstract - Currently there is a variety of designs and Implementation of biometric especially fingerprint. There is currently a standard used for determining matching threshold, which allows vendors to skew their test results in their favour by using assumed figure between -1 to +1 or values between 1 and 100%. The research contribution in this research work is to formulate an equation to determine the threshold against which the minutia matching score will be compare using the features set of the finger itself which is devoid of assumptions. Based on the results of this research, it shows that the proposed design and development of a fingerprint-based identity verification system can be achieved without riding on assumptions. Thereby, eliminating the false rate of Acceptance and reduce false rate of rejection as a result of the threshold computation using the features of the enrolled finger. Further research can be carried out in the area of comparing matching result generated from the threshold assumption with the threshold computation formulated in this thesis paper.

Keywords: Biometrics; Threshold; Matching; Algorithm; Scoring.

16. Paper 29021682: A Sliding Mode Controller for Urea Plant (pp. 115-126)

M. M. Saafan, M. M. Abdelsalam, M. S. Elksasy, S. F. Saraya, and F. F.G. Areed
Computers and Control Systems Engineering Department, Faculty of Engineering, Mansoura University, Egypt.

Abstract - The present paper introduces the mathematical model of urea plant and suggests two methods for designing special purpose controllers. The first proposed method is PID controller and the second is sliding mode controller (SMC). These controllers are applied for a multivariable nonlinear system as a Urea Reactor system. The main target of the designed controllers is to reduce the disturbance of NH3 pump and CO2 compressor in order to reduce the pollution effect in such chemical plant. Simulation results of the suggested PID controller are compared with that of the SMC controller. Comparative analysis proves the effectiveness of the suggested SMC controller than the PID controller according to disturbance minimization as well as dynamic response. Also, the paper presents the results of applying SMC, while maximizing the production of the urea by maximizing the NH3 flow rate. This controller kept the reactor temperature, the reactor pressure, and NH3/CO2 ratio in the suitable operating range. Moreover, the suggested SMC when compared with other controllers in the literature shows great success in maximizing the production of urea.

Keywords: Sliding mode controller, PID controller, urea reactor, Process Control, Chemical Industry, Adaptive controller, Nonlinearity.

17. Paper 29021683: Transmission Control Protocol and Congestion Control: A Review of TCP Variants (pp. 127-135)

Babatunde O. Olasoji, Oyenike Mary Olanrewaju, Isaiah O. Adebayo
Mathematical Sciences and Information Technology Department, Federal University Dutsinma, Katsina State, Nigeria.

Abstract - Transmission control protocol (TCP) provides a reliable data transfer in all end-to-end data stream services on the internet. There are some mechanisms that TCP has that make it suitable for this purpose. Over the years, there have been modifications in TCP algorithms starting from the basic TCP that has only slow-start and congestion avoidance algorithm to the modifications and additions of new algorithms. Today, TCP comes in various variants which include TCP Tahoe, Reno, new Reno, Vegas, sack etc. Each of this TCP variant has its peculiarities, merits and demerits. This paper is a review of four TCP variants, they are: TCP Tahoe, Reno, new Reno and Vegas, their congestion avoidance algorithms, and possible future research areas.

Keywords – Transmission control protocol; Congestion Control; TCP Tahoe; TCP Reno; TCP New Reno; TCP Vegas

18. Paper 31011656: Detection of Black Hole Attacks in MANETs by Using Proximity Set Method (pp. 136-145)

K. Vijaya Kumar, Research Scholar (Karpagam University), Assistant Professor, Department of Computer Science Engineering, Vignan's Institute of Engineering for Women, Visakhapatnam, Andhra Pradesh, India.
Dr. K. Somasundaram, Professor, Department of Computer Science and Engg., Vel Tech High Tech Dr.RR Dr.SR Engineering College, Avadi, Chennai,Tamilnadu India

Abstract - A Mobile Adhoc Networks (MANETS) is an infrastructure less or self-configuring network which contain a collection of mobile nodes moving randomly by changing their topology with limited resources. These Networks are prone to different types of attacks due to lack of central monitoring facility. The main aim is to inspect the effect of black hole attack on the network layer of MANET. A black hole attack is a network layer attack also called sequence number attack which utilizes the destination sequence number to claim that it has a shortest route to reach the destination and consumes all the packets forwarded by the source. To diminish the effects of such attack, we have proposed a detection technique by using Proximity Set Method (PSM) that efficiently detects the malicious nodes in the network. The severity of attack depends on the position of the malicious node that is near, midway or far from the source. The various network scenarios of MANETS with AODV routing protocol are simulated using NS2 simulator to analyze the performance with and without the black hole attack. The performance parameters like PDR, delay, throughput, packet drop and energy consumption are measured. The overall throughput and PDR increases with the number of flows but reduces with the attack. With the increase in the black hole attackers, the PDR and throughput reduces and close to zero as the number of black hole nodes are maximum. The packet drop also increases with the attack. The overall delay factor varies based on the position of the attackers. As the mobility varies the delay and packet drop increases but PDR and throughput decreases as the nodes moves randomly in all directions. Finally the simulation results gives a very good comparison of performance of MANETS with original AODV, with black hole attack and applying proximity set method for presence of black hole nodes different network scenarios.

Keywords: AODV protocol, security, black hole attack, NS2 simulator, proximity set method, performance parameters.

19. Paper 290216995: A Greedy Approach to Out-Door WLAN Coverage Planning (pp. 146-152)

Gilbert M. Gilbert, College of Informatics and Virtual Education, The University of Dodoma

Abstract — Planning for optimal out-door wireless network coverage is one of the core issues in network design. This paper considers coverage problem in outdoor-wireless networks design with the main objective of proposing methods that offer near-optimal coverage. The study makes use of the greedy algorithms and some specified criteria (field strength) to find minimum number of base stations and access points that can be activated to provide maximum services (coverage) to a specified number of users. Various wireless network coverage planning scenarios were considered to an imaginary town subdivided into areas and a comprehensive comparison among them was done to offer desired network coverage that meet the objective.

Keywords — greedy algorithms, outdoor-wlan, coverage planning, greedy algorithms, path loss.

20. Paper 29021620: Cerebellar Model Articulation Controller Network for Segmentation of Computer Tomography Lung Image (pp. 153-157)

(1) Benita K.J. Veronica, (2) Purushothaman S., Rajeswari P.,
(1) Mother Teresa Women's University, Kodaikanal, India.
(2) Associate Professor, Institute of Technology, Haramaya University, Ethiopia.

Abstract - This paper presents the implementation of CMAC network for segmentation of computed tomography lung slice. Representative features are extracted from the slice to train the CMAC algorithm. At the end of training, the final weights are stored in the database. During the testing the CMAC, a lung slice is presented to obtain the segmented image.

Keywords: CMAC; segmentation; computed tomography; lung slice

21. Paper 29021625: Performance Evaluation of Pilot-Aided Channel Estimation for MIMO-OFDM Systems (pp. 158-162)

B. Soma Sekhar, Dept of ECE, Sanketika Vidhya Parishad Engg. College, Visakhapatnam, Andhra Pradesh, India
A. Mallikarjuna Prasad, Dept of ECE, University College of Engineering, Kakinada, JNTUK, Andhra Pradesh, India

Abstract — In this paper a pilot aided channel estimation for Multiple-Input Multiple-Output/Orthogonal Frequency-Division Multiplexing (MIMO/ OFDM) systems in time-varying wireless channels is considered. Channel coefficients can be modeled by using truncated discrete Fourier Basis Expansion model (Fourier-BEM) and a discrete prolate spheroidal sequence model (DPSS). The channel is assumed which is varying linearly with respect to time. Based on these models, a weighted average approach is adopted for estimating LTV channels for OFDM symbols. The performance analysis between Fourier BEM, DPSS models, Legendre and Chebishev polynomial based on Mean square error (MSE) is present. Simulation results show that the DPSS-BEM model outperforms the Fourier Basis expansion model.

Index Terms — Basis Expansion Model (BEM), Discrete Prolate Spheroidal Sequence (DPSS), Mean Square Error (MSE).

22. Paper 29021674: Investigating the Distributed Load Balancing Approach for OTIS-Star Topology (pp. 163-171)

Ahmad M. Awwad, Jehad Al-Sadi

Abstract — This research effort investigates and proposes an efficient method for load balancing problem for the OTIS-Star topology. The proposed method is named OTIS-Star Electronic-Optical-Electronic Exchange Method; OSEOEM; which utilizes the electronic and optical technologies facilitated by the OTIS-Star topology. This method is based on the previous FOFEM algorithm for OTIS-Cube networks. A complete investigation of the OSEOEM is introduced in this paper including a description of the algorithm and the stages of performing Load Balancing, A comprehensive analytical and theoretical study to prove the efficiency of this method, and statistical outcomes based on common used performance measures has been also presented. The outcome of this investigation proves the efficiency of the proposed OSEOEM method.

Keywords — Electronic Interconnection Networks, Optical Networks, Load balancing, Parallel Algorithms, OTIS-Star Network.

23. Paper 29021694: American Sign Language Pattern Recognition Based on Dynamic Bayesian Network (pp. 172-177)

Habes Alkhraisat, Saqer Alshrah
Department of Computer Science, Al-Balqa Applied University, Jordan

Abstract — Sign languages are usually developed among deaf communities, which include friends and families of deaf people or people with hearing impairment. American Sign Language (ASL) is the primary language used by the American Deaf Community. It is not simply a signed representation of English, but rather, a rich natural language with a unique structure, vocabulary, and grammar. In this paper, we propose a method for American Sign Language alphabet, and number gestures interpretation in a continuous video stream using a dynamic Bayesian network. The experimental result, using RWTHBOSTON-104 data set, shows a recognition rate upwards of 99.09%.

Index Terms — American Sign Language (ASL), Dynamic Bayesian Network, Hand Tracking, Feature extraction.

24. Paper 2902169910: Identification of Breast Cancer by Artificial Bee Colony Algorithm with Least Square Support Vector Machine (pp. 178-183)

S. Mythili, PG & Research Department of Computer Application, Hindusthan College of Arts & Science, Coimbatore, India
Dr. A. V. Senthilkumar, Director, PG & Research Department of Computer Application, Hindusthan College of Arts & Science, Coimbatore, India

Abstract - Procedure for the identification of several discriminant factors. A new method is proposed for identification of Breast Cancer in Peripheral Blood with microarray Datasets by introducing the Hybrid Artificial Bee Colony (ABC) algorithm with Least Squares Support Vector Machine (LS-SVM), namely as ABC-SVM. Breast cancer is identified by Circulating Tumor Cells in the Peripheral Blood. The mechanisms that implicate Circulating Tumor Cells (CTC) in metastatic disease is notably in Metastatic Breast Cancer (MBC), remain elusive. The proposed work is focused on the identification of tissues in Peripheral Blood that can indirectly reveal the presence of cancer cells. By selecting publicly available Breast Cancer tissues and Peripheral Blood microarray datasets, we follow two-step elimination.

Keywords: Breast Cancer (BC), Circulating Tumor Cells (CTC), Peripheral Blood (PB), Artificial Bee Colony (ABC), Least Squares Support Vector Machine (LSSVM).

25. Paper 29021601: Moving Object Segmentation and Vibrant Background Elimination Using LS-SVM (pp. 184-197)

Mehul C. Parikh, Computer Engineering Department, Chartoar University of Science and Technology, Changa, Gujarat, India.
Kishor G. Maradia, Department ofElectronics and Communication, Government Engineering College, Gandhinagar, Gujarat, India

Abstract - Moving object segmentation is a significant research area in the field of computer intelligence due to technological and theoretical progress. Many approaches are being developed for moving object segmentation. These approaches are useful for specific situation but have many restrictions. Execution speed of these approaches is one of the major limitations. Machine learning techniques are used to decrease time and improve quality of result. LS-SVM optimizes result quality and time complexity in classification problem. This paper describes an approach to segment moving object and vibrant background elimination using the least squares support vector machine method. In this method consecutive frame difference was given as an input to bank of Gabor filter to detect texture feature using pixel intensity. Mean value of intensity on 4 * 4 block of image and on whole image was calculated and which are then used to train LS-SVM model using random sampling. Trained LS-SVM model was then used to segment moving object from the image other than the training images. Results obtained by this approach are very promising with improvement in execution time.

Key Words: Segmentation, Machine Learning, Gabor filter, LS-SVM.

26. Paper 29021609: On Annotation of Video Content for Multimedia Retrieval and Sharing (pp. 198-218)

Mumtaz Khan, Shah Khusro, Irfan Ullah
Department of Computer Science, University of Peshawar, Peshawar 25120, Pakistan

Abstract - The development of standards like MPEG-7, MPEG-21 and ID3 tags in MP3 have been recognized from quite some time. It is of great importance in adding descriptions to multimedia content for better organization and retrieval. However, these standards are only suitable for closed-world-multimedia-content where a lot of effort is put in the production stage. Video content on the Web, on the contrary, is of arbitrary nature captured and uploaded in a variety of formats with main aim of sharing quickly and with ease. The advent of Web 2.0 has resulted in the wide availability of different video-sharing applications such as YouTube which have made video as major content on the

Web. These web applications not only allow users to browse and search multimedia content but also add comments and annotations that provide an opportunity to store the miscellaneous information and thought-provoking statements from users all over the world. However, these annotations have not been exploited to their fullest for the purpose of searching and retrieval. Video indexing, retrieval, ranking and recommendations will become more efficient by making these annotations machine-processable. Moreover, associating annotations with a specific region or temporal duration of a video will result in fast retrieval of required video scene. This paper investigates state-of-the-art desktop and Web-based-multimedia annotation-systems focusing on their distinct characteristics, strengths and limitations. Different annotation frameworks, annotation models and multimedia ontologies are also evaluated.

Keywords: Ontology, Annotation, Video sharing web application

27. Paper 29021617: A New Approach for Energy Efficient Linear Cluster Handling Protocol In WSN (pp. 219-227)

Jaspinder Kaur, Varsha Sahni

Abstract - Wireless Sensor Networks (WSN) is a rising field for researchers in the recent years. For obtaining durability of network lifetime, and reducing energy consumption, energy efficiency routing protocol play an important role. In this paper, we present an innovative and energy efficient routing protocol. A New linear cluster handling (LCH) technique towards Energy Efficiency in Linear WSNs with multiple static sinks [4] in a linearly enhanced field of 1500m*350m2. We are divided the whole into four equal sub-regions. For efficient data gathering, we place three static sinks i.e. one at the centre and two at the both corners of the field. A reactive and Distance plus energy dependent clustering protocol Threshold Sensitive Energy efficient with Linear Cluster Handling [4] DE (TEEN-LCH) is implemented in the network field. Simulation shows improved results for our proposed protocol as compared to TEEN-LCH, in term of throughput, packet delivery ratio and energy consumption.

Keywords: WSN; Routing Protocol; Throughput; Energy Consumption; Packet Delivery

28. Paper 29021624: Protection against Phishing in Mobile Phones (pp. 228-233)

Avinash Shende, IT Department, SRM University, Kattankulathur, Chennai, India
Prof. D. Saveetha, Assistant Professor, IT Department, SRM University, Kattankulathur, Chennai, India

Abstract - Phishing is the attempt to get confidential information such as user-names, credit card details, passwords and pins, often for malicious reasons, by making people believe that they are communicating with legitimate person or identity. In recent years we have seen increase in threat of phishing on mobile phones. In fact, mobile phone phishing is more dangerous than phishing on desktop because of limitations of mobile phones like mobile user habits and small screen. Existing mechanism made for detecting phishing attacks on computers are not able to avoid phishing attacks on mobile devices. We present an anti-phishing mechanism for mobile devices. Our solution verifies if webpages is legitimate or not by comparing the actual identity of webpage with the claimed identity of the webpage. We will use OCR tool to find the identity claimed by the webpage.

29. Paper 29021626: Hybrid Cryptography Technique for Information Systems (pp. 234-243)

Zohair Malki
Faculty of Computer Science and Engineering, Taibah University, Yanbu, Saudi Arabia

Abstract - Information systems based applications are increasing rapidly in many fields including educational, medical, commercial and military areas, which have posed many security and privacy challenges. The key component of any security solution is encryption. Encryption is used to hide the original message or information in a new form that can be retrieved by the authorized users only. Cryptosystems can be divided into two main types: symmetric and asymmetric systems. In this paper we discussed some common systems that belong to both types. Specifically, we will discuss, compare and test the implementation for RSA, RC5, DES, Blowfish and Twofish. Then, a new hybrid

system composed of RSA and RC5 is proposed and tested against these two systems when each used alone. The obtained results show that the proposed system achieves better performance.

30. Paper 29021629: An Efficient Network Traffic Filtering that Recognize Anomalies with Minimum Error Received (pp. 244-256)

Mohammed N. Abdul Wahid and Azizol Bin Abdullah
Department of Communication Technology and Networks, Faculty of Computer Science and Information Technology, University Putra Malaysia, Malaysia

Abstract - The main method is related to processing and filtering data packets on a network system and, more specifically, analyzing data packets transmitted on a regular speed communications links for errors and attackers' detection and signal integrity analysis. The idea of this research is to use flexible packet filtering which is a combination of both the static and dynamic packet filtering with the margin of support vector machine. Many experiments have been conducted in order to investigate the performance of the proposed schemes and comparing them with recent software's that is most relatively to our proposed method that measuring the bandwidth, time, speed and errors. These experiments are performed and examined under different network environments and circumstances. The comparison has been done and results proved that our method gives less error received from the total analyzed packets.

Keywords: Anomaly Detection, Data Mining, Data Processing, Flexible Packet Filtering, Misuse Detection, Network Traffic Analyzer, Packet sniffer, Support Vector Machine, Traffic Signature Matching, User Profile Filter.

31. Paper 29021634: Proxy Blind Signcryption Based on Elliptic Curve Discrete Logarithm Problem (pp. 257-262)

Anwar Sadat, Department of Information Technology, kohat University of Science and Technology K-P, Pakistan
Insaf Ullah, Hizbullah Khattak, Sultan Ullah, Amjad-ur-Rehman
Department of Information Technology, Hazara Uuniversity Mansehra K-P, Pakistan

Abstract - Nowadays anonymity, rights delegations and hiding information play primary role in communications through internet. We proposed a proxy blind signcryption scheme based on elliptic curve discrete logarithm problem (ECDLP) meet all the above requirements. The design scheme is efficient and secure because of elliptic curve crypto system. It meets the security requirements like confidentiality, Message Integrity, Sender public verifiability, Warrant unforgeability, Message Unforgeability, Message Authentication, Proxy Non-Repudiation and blindness. The proposed scheme is best suitable for the devices used in constrained environment.

Keywords: proxy signature, blind signature, elliptic curve, proxy blind signcryption.

32. Paper 29021646: A Comprehensive Survey on Hardware/Software Partitioning Process in Co-Design (pp. 263-279)

Imene Mhadhbi, Slim BEN OTHMAN, Slim Ben Saoud
Department of Electrical Engineering, National Institute of Applied Sciences and Technology, Polytechnic School of Tunisia, Advanced Systems Laboratory, B.P. 676, 1080 Tunis Cedex, Tunisia

Abstract - Co-design methodology deals with the problem of designing complex embedded systems, where Hardware/software partitioning is one key challenge. It decides strategically the system's tasks that will be executed on general purpose units and the ones implemented on dedicated hardware units, based on a set of constraints. Many relevant studies and contributions about the automation techniques of the partitioning step exist. In this work, we explore the concept of the hardware/software partitioning process. We also provide an overview about the historical achievements and highlight the future research directions of this co-design process.

Keywords: Co-design; embedded system; hardware/software partitioning; embedded architecture

33. Paper 29021647: Heterogeneous Embedded Network Evaluation of CAN-Switched ETHERNET Architecture (pp. 280-294)

Nejla Rejeb, Ahmed Karim Ben Salem, Slim Ben Saoud
LSA Laboratory, INSAT-EPT, University of Carthage, TUNISIA

Abstract - The modern communication architecture of new generation transportation systems is described as heterogeneous. This new architecture is composed by a high rate Switched ETHERNET backbone and low rate data peripheral buses coupled with switches and gateways. Indeed, Ethernet is perceived as the future network standard for distributed control applications in many different industries: automotive, avionics and industrial automation. It offers higher performance and flexibility over usual control bus systems such as CAN and Flexray. The bridging strategy implemented at the interconnection devices (gateways) presents a key issue in such architecture. The aim of this work consists on the analysis of the previous mixed architecture. This paper presents a simulation of CAN-Switched Ethernet network based on OMNET++. To simulate this network, we have also developed a CAN-Switched Ethernet Gateway simulation model. To analyze the performance of our model we have measured the communication latencies per device and we have focused on the timing impact introduced by various CAN-Ethernet multiplexing strategies at the gateways. The results herein prove that regulating the gateways CAN remote traffic has an impact on the end to end delays of CAN flow. Additionally, we demonstrate that the transmission of CAN data over an Ethernet backbone depends heavily on the way this data is multiplexed into Ethernet frames.

Keywords: Ethernet, CAN, Heterogeneous Embedded networks, Gateway, Simulation, End to end delay.

34. Paper 29021660: Reusability Quality Attributes and Metrics of SaaS from Perspective of Business and Provider (pp. 295-312)

Areeg Samir, Nagy Ramadan Darwish
Information Technology and System, Institute of Statistical Studies and Research

Abstract - Software as a Service (SaaS) is defined as a software delivered as a service. SaaS can be seen as a complex solution, aiming at satisfying tenants requirements during runtime. Such requirements can be achieved by providing a modifiable and reusable SaaS to fulfill different needs of tenants. The success of a solution not only depends on how good it achieves the requirements of users but also on modifies and reuses provider's services. Thus, providing reusable SaaS, identifying the effectiveness of reusability and specifying the imprint of customization on the reusability of application still need more enhancements. To tackle these concerns, this paper explores the common SaaS reusability quality attributes and extracts the critical SaaS reusability attributes based on provider side and business value. Moreover, it identifies a set of metrics to each critical quality attribute of SaaS reusability. Critical attributes and their measurements are presented to be a guideline for providers and to emphasize the business side.

Index Terms - Software as a Service (SaaS), Quality of Service (QoS), Quality attributes, Metrics, Reusability, Customization, Critical attributes, Business, Provider.

35. Paper 29021661: A Model Driven Regression Testing Pattern for Enhancing Agile Release Management (pp. 313-333)

Maryam Nooraei Abadeh, Department of Computer Science, Science and Research Branch, Islamic Azad University, Tehran, Iran
Seyed-Hassan Mirian-Hosseinabadi, Department of Computer Science, Sharif University of Technology, Tehran, Iran

Abstract - Evolutionary software development disciplines, such as Agile Development (AD), are test-centered, and their application in model-based frameworks requires model support for test development. These tests must be applied against changes during software evolution. Traditionally regression testing exposes the scalability problem, not only in terms of the size of test suites, but also in terms of complexity of the formulating modifications and keeping the fault detection after system evolution. Model Driven Development (MDD) has promised to reduce the complexity of software maintenance activities using the traceable change management and automatic change propagation. In this paper, we propose a formal framework in the context of agile/lightweight MDD to define generic test models, which can be automatically transformed into executable tests for particular testing template models using incremental model transformations. It encourages a rapid and flexible response to change for agile testing foundation. We also introduce on the-fly agile testing metrics which examine the adequacy of the changed requirement coverage using a new measurable coverage pattern. The Z notation is used for the formal definition of the framework. Finally, to evaluate different aspects of the proposed framework an analysis plan is provided using two experimental case studies.

Keywords: Agile development, Model Driven testing. On-the fly Regression Testing. Model Transformation. Test Case Selection.

36. Paper 29021681: Comparative Analysis of Early Detection of DDoS Attack and PPS Scheme against DDoS Attack in WSN (pp. 334-342)

Kanchan Kaushal, Varsha Sahni
Department of Computer Science Engineering, CTIEMT Shahpur Jalandhar, India

Abstract- Wireless Sensor Networks carry out has great significance in many applications, such as battlefields surveillance, patient health monitoring, traffic control, home automation, environmental observation and building intrusion surveillance. Since WSNs communicate by using radio frequencies therefore the risk of interference is more than with wired networks. If the message to be passed is not in an encrypted form, or is encrypted by using a weak algorithm, the attacker can read it, and it is the compromise to the confidentiality. In this paper we describe the DoS and DDoS attacks in WSNs. Most of the schemes are available for the detection of DDoS attacks in WSNs. But these schemes prevent the attack after the attack has been completely launched which leads to data loss and consumes resources of sensor nodes which are very limited. In this paper a new scheme early detection of DDoS attack in WSN has been introduced for the detection of DDoS attack. It will detect the attack on early stages so that data loss can be prevented and more energy can be reserved after the prevention of attacks. Performance of this scheme has been seen by comparing the technique with the existing profile based protection scheme (PPS) against DDoS attack in WSN on the basis of throughput, packet delivery ratio, number of packets flooded and remaining energy of the network.

Keywords: DoS and DDoS attacks, Network security, WSN

37. Paper 29021687: Detection of Stealthy Denial of Service (S-DoS) Attacks in Wireless Sensor Networks (pp. 343-348)

Ram Pradheep Manohar, St. Peter's University, Chennai
E. Baburaj, Narayanaguru College of Engineering, Nagercoil

Abstract — Wireless sensor networks (WSNs) supports and involving various security applications like industrial automation, medical monitoring, homeland security and a variety of military applications. More researches highlight the need of better security for these networks. The new networking protocols account the limited resources available in WSN platforms, but they must tailor security mechanisms to such resource constraints. The existing denial of service (DoS) attacks aims as service denial to targeted legitimate node(s). In particular, this paper address the stealthy denial-of-service (S-DoS)attack, which targets at minimizing their visibility, and at the same time, they can be as harmful as other attacks in resource usage of the wireless sensor networks. The impacts of Stealthy Denial of Service (S-DoS) attacks involve not only the denial of the service, but also the resource maintenance costs in terms of resource usage. Specifically, the longer the detection latency is, the higher the costs to be incurred. Therefore, a particular attention has to be paid for stealthy DoS attacks in WSN. In this paper, we propose a new attack strategy namely Slowly Increasing and Decreasing under Constraint DoS Attack Strategy (SIDCAS) that leverage the application vulnerabilities, in order to degrade the performance of the base station in WSN. Finally we analyses the characteristics of the S-DoS attack against the existing Intrusion Detection System (IDS) running in the base station.

Index Terms— resource constraints, denial-of-service attack, Intrusion Detection System

38. Paper 2902169912: Intelligent Radios in the Sea (pp. 349-357)

Ebin K. Thomas
Amrita Center for Wireless Networks & Applications (AmritaWNA), Amrita School of Engineering, Amritapuri, Amrita Vishwa Vidyapeetham, Amrita University, India

Abstract - Communication over the sea has huge importance due to fishing and worldwide trade transportation. Current communication systems around the world are either expensive or use dedicated spectrum, which lead to crowded spectrum usage and eventually low data rates. On the other hand, unused frequency bands of varying bandwidths within the licensed spectrum have led to the development of new radios termed Cognitive radios that can intelligently capture the unused bands opportunistically by sensing the spectrum. In a maritime network where data of different bandwidths need to be sent, such radios could be used for adapting to different data rates. However, there is not much research conducted in implementing cognitive radios to maritime environments. This exploratory article introduces the concept of cognitive radio, the maritime environment, its requirements and surveys, and some of the existing cognitive radio systems applied to maritime environments.

Keywords — Cognitive Radio, Maritime Network, Spectrum Sensing.

39. Paper 290216991: Developing Context Ontology using Information Extraction (pp. 358-363)

*Ram kumar #, Shailesh Jaloree #, R S Thakur **
 # Applied Mathematics & Computer Application, SATI, Vidisha, India
** MANIT, Bhopal, India*

Abstract — Information Extraction addresses the intelligent access to document contents by automatically extracting information applicable to a given task. This paper focuses on how ontologies can be exploited to interpret the contextual document content for IE purposes. It makes use of IE systems from the point of view of IE as a knowledge-based NLP process. It reviews the dissimilar steps of NLP necessary for IE tasks: Rule-Based & Dependency Based Information Extraction, Context Assessment.

40. Paper 31031601: Challenges and Interesting Research Directions in Model Driven Architecture and Data Warehousing: A Survey (pp. 364-398)

Amer Al-Badarneh, Jordan University of Science and Technology
Omran Al-Badarneh, Devoteam, Riyadh, Saudi Arabia

Abstract - Model driven architecture (MDA) is playing a major role in today's system development methodologies. In the last few years, many researchers tried to apply MDA to Data Warehouse Systems (DW). Their focus was on automatic creation of Multidimensional model (Start schema) from Conceptual Models. Furthermore, they addressed the conceptual modeling of QoS parameters such as Security in early stages of system development using MDA concepts. However, there is a room to improve further the DW development using MDA concepts. In this survey we identify critical knowledge gaps in MDA and DWs and make a chart for future research to motivate researchers to close this breach and improve DW solution's quality and performance, and also minimize drawbacks and limitations. We identified promising challenges and potential research areas that need more work on it. Using MDA to handle DW performance, multidimensionality and friendliness aspects, applying MDA to other stages of DW development life cycle such as Extracting, Transformation and Loading (ETL) Stage, developing On Line Analytical Processing(OLAP) end user Application, applying MDA to Spatial and Temporal DWs, developing a complete, self-contained DW framework that handles MDA-technical issues together with managerial issues using Capability Maturity Model Integration(CMMI) standard or International standard Organization (ISO) are parts of our findings.

Keywords: Data warehousing, Model driven Architecture (MDA), Platform Independent Model (PIM), Platform Specific Model (PSM), Common Warehouse Metamodel (CWM), XML Metadata Interchange (XMI).

41. Paper 290216992: A Framework for Building Ontology in Education Domain for Knowledge Representation (pp. 399-403)

Monica Sankat #, R S Thakur, Shailesh Jaloree #*
 # Department of Applied Mathematics and Computer Application, SATI, Vidisha, India
** Department of Computer Application, MANIT, Bhopal, India*

Abstract — In this paper we have proposed a method of creating domain ontology using protégé tool. Existing ontology does not take the semantic into context while displaying the information about different modules. This paper proposed a methodology for the derivation and implementation of ontology in education domain using protégé 4.3.0 tool.

42. Paper 2902169913: Mobility Aware Multihop Clustering based Safety Message Dissemination in Vehicular Ad-hoc Network (pp. 404-417)

Nishu Gupta, Dr. Arun Prakash, Dr. Rajeev Tripathi
Department of Electronics and Communication Engineering
Motilal Nehru National Institute of Technology Allahabad, Allahabad-211004, INDIA

Abstract - A major challenge in Vehicular Ad-hoc Network (VANET) is to ensure real-time and reliable dissemination of safety messages among vehicles within a highly mobile environment. Due to the inherent characteristics of VANET such as high speed, unstable communication link, geographically constrained topology and varying channel capacity, information transfer becomes challenging. In the multihop scenario, building and maintaining a route under such stringent conditions becomes even more challenging. The effectiveness of traffic safety applications using VANET depends on how efficiently the Medium Access Control (MAC) protocol has been designed. The main challenge while designing such a MAC protocol is to achieve reliable delivery of messages within the time limit under highly unpredictable vehicular density. In this paper, Mobility aware Multihop Clustering based Safety message dissemination MAC Protocol (MMCS-MAC) is proposed in order to accomplish high reliability, low communication overhead and real time delivery of safety messages. The proposed MMCS-MAC is capable of establishing a multihop sequence through clustering approach using Time Division Multiple Access mechanism. The protocol is designed for highway scenario that allows better channel utilization, improves network performance and assures fairness among all the vehicles. Simulation results are presented to verify the effectiveness of the proposed scheme and comparisons are made with the existing IEEE 802.11p standard and other existing MAC protocols. The evaluations are performed in terms of multiple metrics and the results demonstrate the superiority of the MMCS-MAC protocol as compared to other existing protocols related to the proposed work.

Keywords- Clustering, Multihop, Safety, TDMA, V2V,VANET.

43. Paper 290216997: Clustering of Hub and Authority Web Documents for Information Retrieval (pp. 418-422)

Kavita Kanathey, Computer Science, Barkatullah University, Bhopal, India
R. S. Thakur, Department of Computer Application, Maulana Azad National Institute of Technology (MANIT), Bhopal, MP, India
Shailesh Jaloree, Department of applied Mathematics, SATI, Vidisha, Bhopal, MP, India

Abstract - Due to the exponential growth of World Wide Web (or simply the Web), finding and ranking of relevant web documents has become an extremely challenging task. When a user tries to retrieve relevant information of high quality from the Web, then ranking of search results of a user query plays an important role. Ranking provides an ordered list of web documents so that users can easily navigate through the search results and find the information content as per their need. In order to rank these web documents, a lot of ranking algorithms (PageRank, HITS, Weight

PageRank) have been proposed based upon many factors like citations analysis, content similarity, annotations etc. However, the ranking mechanism of these algorithms gives user with a set of non-classified web documents according to their query. In this paper, we propose a link-based clustering approach to cluster search results returned from link based web search engine. By filtering some irrelevant pages, our approach classified relevant web pages into most relevant, relevant and irrelevant groups to facilitate users' accessing and browsing. In order to increase relevancy accuracy, K-mean clustering algorithm is used. Preliminary evaluations are conducted to examine its effectiveness. The results show that clustering on web search results through link analysis is promising. This paper also outlines various page ranking algorithms.

Keywords - World Wide Web, search engine, information retrieval, Pagerank, HITS, Weighted Pagerank, link analysis.

44. Paper 29021645: Dorsal Hand Vein Identification (pp. 423-433)

Sarah HACHEMI BENZIANE, Irit, University Paul Sabatier France, Simpa, University of Sciences and Technology of Oran, Mohamed Boudiaf Algérie
Abdelkader BENYETTOU, Simpa, University of Sciences and Technology of Oran

Abstract — In this paper, we present an competent approach for dorsal hand vein features extraction from near infrared images. The physiological features characterize the dorsal venous network of the hand. These networks are single to each individual and can be used as a biometric system for person identification/authentication. An active near infrared method is used for image acquisition. The dorsal hand vein biometric system developed has a main objective and specific targets; to get an electronic signature using a secure signature device. In this paper, we present our signature device with its different aims; respectively: The extraction of the dorsal veins from the images that were acquired through an infrared device. For each identification, we need the representation of the veins in the form of shape descriptors, which are invariant to translation, rotation and scaling; this extracted descriptor vector is the input of the matching step. The optimization decision system settings match the choice of threshold that allows to accept / reject a person, and selection of the most relevant descriptors, to minimize both FAR and FRR errors. The final decision for identification based descriptors selected by the PSO hybrid binary give a FAR =0% and FRR=0% as results.

Keywords - Biometrics, identification, hand vein, OTSU, anisotropic diffusion filter, top & bottom hat transform, BPSO,

45. Paper 290216999: Blind Image Separation Based on Exponentiated Transmuted Weibull Distribution (pp. 423-433)

A. M. Adam, Department of Mathematics, Faculty of Science, Zagazig University, P.O. Box, Zagazig, Egypt
R. M. Farouk, Department of Mathematics, Faculty of Science, Zagazig University, P.O. Box, Zagazig, Egypt
M. E. Abd El-aziz, Department of Mathematics, Faculty of Science, Zagazig University, P.O. Box, Zagazig, Egypt

Abstract - In recent years the processing of blind image separation has been investigated. As a result, a number of feature extraction algorithms for direct application of such image structures have been developed. For example, separation of mixed fingerprints found in any crime scene, in which a mixture of two or more fingerprints may be obtained, for identification, we have to separate them. In this paper, we have proposed a new technique for separating a multiple mixed images based on exponentiated transmuted Weibull distribution. To adaptively estimate the parameters of such score functions, an efficient method based on maximum likelihood and genetic algorithm will be used. We also calculate the accuracy of this proposed distribution and compare the algorithmic performance using the efficient approach with other previous generalized distributions. We find from the numerical results that the proposed distribution has flexibility and an efficient result.

Keywords- Blind image separation, Exponentiated transmuted Weibull distribution, Maximum likelihood, Genetic algorithm, Source separation, FastICA.

Performance Evaluation of Pilot-Aided Channel Estimation for MIMO-OFDM Systems

B.Soma sekhar (1), A.MallikarjunaPrasad (2)
(1) Dept of ECE, Sanketika Vidhya Parishad Engg. College, Visakhapatnam, Andhra Pradesh, India

(2)Dept of ECE, University College of Engineering, Kakinada, JNTUK, Andhra Pradesh, India

***Abstract*—In this paper a pilot aided channel estimation for Multiple-Input Multiple-Output/Orthogonal Frequency-Division Multiplexing (MIMO/ OFDM) systems in time-varying wireless channels is considered. Channel coefficients can be modeled by using truncated discrete Fourier Basis Expansion model (Fourier-BEM) and a discrete prolate spheroidal sequence model (DPSS). The channel is assumed which is varying linearly with respect to time. Based on these models, a weighted average approach is adopted for estimating LTV channels for OFDM symbols. The performance analysis between Fourier BEM, DPSS models, Legendre and chebishev polynomial based on Mean square error (MSE) is present. Simulation results show that the DPSS-BEM model outperforms the Fourier Basis expansion model.**

***Index Terms*—Basis Expansion Model (BEM), Discrete Prolate Spheroidal Sequence (DPSS), Mean Square Error (MSE).**

I. INTRODUCTION

THE increasing demand for high-speed reliable wireless communications over the limited radio frequency spectrum has spurred increasing interest in multiple-input multiple-output (MIMO) systems to achieve higher transmission rates. The combination of MIMO and OFDM can achieve a lower error rate and enable high-capacity wireless communication systems. Such systems, however, rely upon the knowledge of channel state information (CSI) which is often obtained through channel estimation.

Fourier basis are used in fading channels when multipath propagation is caused by few dominant reflectors gives rise to linearly varying path delays [1]. In pilot tone channel estimation scheme MIMO-OFDM channel can be estimated by using LS algorithm. The drawback of OFDM system based on pilot tones is that the transmit antennas require more pilot tones for training which results in reduction of efficiency [2]. One of the suitable techniques that can be applied to the modeling of a time variant frequency selective channel is slepian basis expansion. Numerical complexity (with same number of unknown) of Slepian basis expansion is 3 magnitudes smaller when compared with Fourier basis expansion [3]. By using Superimposed pilot time domain channel estimation is developed for fast varying OFDM channels. Temporary channel estimates can be found by resorting LS channel estimator using pilot symbols [4]. An alternative approach for estimating the LTV channels of MIMO-OFDM systems using superimposed training has been studied. In Superimposed training LTV channels are modeled by truncated DFB and then a two-step approach was investigated [5]. The simulations offered in this paper shows the comparison results of Fourier basis model and DPSS BEM model Legendre polynomial and Chebyshev Polynomial**.** The remaining sections of the paper are planned as follows. Section II presents the MIMO-OFDM system model. The Fourier BEM, DPS sequences Legendre polynomial and Chebyshev polynomials are defined in Section III. The Analysis of channel estimation was defined in Section IV. Section V presents simulations and results. We conclude the paper with Section VI.

II. MIMO-OFDM SYSTEM MODEL

The MIMO-OFDM system of N transmit and M receive antennas are considered [5]. We use the IFFT at the transmitter side for which the modulated output can be expressed as

$$X_n(i) = [x_n(i,0),..., x_n(i,t),..., x_n(i,B-1)]^T \quad (1)$$

Where n=1,2,…. ,N

$X_n(i)$ is concatenated by the cyclic prefix (CP) of length $\overline{L}$ (to cancel the inter-symbol interference (ISI) $\overline{L}$ should be larger or at least equal to the maximum channel delay L), that must be propagated through the respective channels. The signals received at m_{th} receive antenna removes the cyclic prefix (CP) and then piles the received signals $y^{(m)}(i,t)$ t=0,…, B-1. This can be written in a vector form as

$$Y^{(m)}(x) = [y^{(m)}(i,0),..., y^{(m)}(i,t),..., y^{(m)}(i,B-1)]^T,$$

m=1,…,M (2)

and the received signals $y^{(m)}(i,t)$ in (2) is given by

$$y^{(m)}(i,t) = \sum_{n=1}^{N} X_n(i) \otimes h_n^{(m)} + v^{(m)}(i,t) \quad (3)$$

Where $h_n^{(m)}(t) = [h_{n,0}^{(m)}(t),..., h_{n,L-1}^{(m)}(t), 0_{1\times B-L}]^T$ is the impulse response vector of the propagating channel from the nth transmit to the mth receive antenna. The coefficients of the channel $h_{n,l}^{(m)}(t)$, l=0,…,L-1 are time variable functions and $v^{(m)}(i,t)$ is the additive Gaussian noise.

III. BASIS EXPANSION MODELS

In Wireless communication, the variations of channel with respect to time arise due to mobility nature of the transmitter and the receiver and with multipath effects [8]. We confine this type of variations using statistical models. In wireless applications the fading channels can be represented using basis expansion models [1], [7]. The different types of BEMs are Fourier basis functions [1], [9], Discrete Prolate Spheroidal sequences [6], [10] polynomial basis [7], universal expansion model, probability expansion model. In this paper we are studying Fourier Basis Expansion model, Legendre Polynomial, Chebeshev Polynomial and Discrete Prolate Spheroidal sequences (DPSS).

A. Fourier Basis Expansion Model

The approximation of the channel coefficients in (3) using Discrete Fourier basis within the time is

$$h_{n,l}^{(m)}(t)=\sum_{q=0}^{Q-1} h_{n,l,q}^{(m)} e^{\frac{-j2\pi(q-Q/2)t}{\Omega}} \tag{4}$$

t = (l-1) Ω,……..,lΩ , l=1,2,3,…

t represents the time interval and $h_{n,l,q}^{(m)}$ is a constant coefficient, the order of the basis expansion is Q and is described as Q $\geq 2f_d\,\Omega/f_s$ [1], [4]. The length of segment is Ω > B with segment index l. The frame has more number of OFDM symbols, which are denoted by i = 1, ⋯ , I where I =Ω/ B' and $B'=B+\overline{L}$. At receiver, an FFT operation is performed on the vector (2), and the demodulated outputs can be written as

$$U^{(m)}(i)=[u^{(m)}(i,0),...,u^{(m)}(i,k),...,u^{(m)}(i,B-1)]^T$$

$$=FY^{(m)}(i) \text{ ,m=1,…,M} \tag{5}$$

Using (3), we can write the FFT demodulated signals in (4) as

$$u^{(m)}(i,t)=FFT\{\sum_{n=1}^{N}\sum_{l=0}^{L-1} h_{n,l}^{(m)}(t)x_n(i,t-1)+v^{(m)}(i,t)\}$$

$$=\sum\nolimits_{n=1}^{N}\sum\nolimits_{l=0}^{L-1} FFT\{h_{n,l}^{(m)}(t)\}\otimes FFT\{x_n(i,t)\}+v^{-(m)}(i,k)$$

$$=\sum\nolimits_{n=1}^{N}\sum\nolimits_{l=0}^{L-1} FFT\{h_{n,lq}^{(m)} e^{\frac{-j2\pi(q-Q/2)}{\Omega}}\}\otimes S_n(i)+v^{-(m)}(i,k)$$

(6)

Where FFT {•} represents the FFT vector of the specified function, $v^{-(m)}(i,k)$ is the additive noise in frequency-domain. If the channel changes slowly when compared to the duration of an OFDM symbol then, the channel variations and the resulting ICI can be neglected.

B. Legendre Polynomial

Legendre basis expansion model (LBEM) has been used in modeling the fading channels. In this paper, Legendre polynomials are used as basis expansion model for doubly-selective fading channels. The Legendre polynomials are the solution to the following differential equation

$$h_{n,l}^{(m)}(t)=\sum_{q=0}^{Q=1} h_{n,l,q}^{(m)} P_i(x) \tag{7}$$

i = {0,1,2…}

Where $P_i(x)$ is the Legendre polynomial of order i. By having $P_0(x)$ = 1 and $P_1(x)$ = x, the Bonnet recursion formula is given by

$$P_i(x)=\frac{1}{2^i i!}\frac{d^i}{dx^i}\left[(x^2-1)^i\right] \tag{8}$$

The property of Legendre Polynomials is that they are orthogonal on $-1\leq x\leq 1$ interval as below.

$$\int_{-1}^{1} P_n(x)P_m(x)dx=\frac{2}{2i+1}\delta_{nm} \tag{9}$$

Where δ_{nm} is the Kronecker delta.

C. Chebyshev Polynomial

The approximation of Chebyshev coefficients given by the Chebyshev polynomials of the first kind can be developed by means of generating function

$$h_{n,l}^{(m)}(t)=\sum_{q=0}^{Q=1} h_{n,l,q}^{(m)} T_i(x) \tag{10}$$

Where
i = {0,1,2…} of $T_i(x)$ is the Chebyshev polynomial of order i . By having $T_0(x)$ = 1 and $T_1(x)$ = x using the recursion formula is given by

$$T_i(x)=\frac{(-2)^i i!}{(2i)!}\sqrt{(1-x^2)}\,\frac{d^i}{dx^i}\left[(1-x^2)^{i-1/2}\right] \tag{11}$$

The property of Chebyshev Polynomials is that they are orthogonal on $-1\leq x\leq 1$ interval with respect to the weight $\sqrt{x^2-1}$ as given below.

$$\int_{-1}^{1}\frac{T_m(x)T_n(x)dx}{\sqrt{1-x^2}} \tag{12}$$

D. Discrete Prolate Spheroidal Sequence (DPSS) model

DPSS modal represented in terms of band limited sequence with Q number of basis function is implemented in order to avoid the insufficiency of fourier basis expansion modal as it requires a low-dimensional sub-space which is orthogonal to each other. This achieves the drawback (windowing) occurred in the Fourier basis expansion [1] while enabling the parameter estimation**.** The slepian basis functions are band-limited to the known maximum variation of channel time.

For time-varying channels a parsimonious representation is provided by BEMs, where one can assume that

$$h_{n,l}^{(m)}(t) = \sum_{q=0}^{Q-1} h_{n,l,q}^{(m)} \mu_q(t) \text{ ,t=0,1,...,}\Omega \tag{13}$$

Where $\mu_q(.)$ is the scalar q[th] basis function (q =1,..., Q). For each block these basis functions $\mu_q(t)$ are common to all users. Consider the continuous varying time channel having a delay spread τ_d sec and a Doppler frequency of f_d Hz. The DPS sequences are orthonormal over the finite time interval t =0, 1,..., Ω.

The Slepian sequences are rectangular windowed versions of many number of DPS sequences that are exactly band limited to the frequency range $[-f_d T_{s,} f_d T_s]$. DPS-BEM [6] outperforms other commonly used BEMs (such as CE-BEM [3], [9] and polynomial BEM) in approximating a Jakes' channel over a wide range of Doppler spreads for the same number of parameters.

The Slepian sequence $\mu_0(t)$ is the single sequence which is band-limited and most of the time in a given time interval and the next sequence $\mu_1(t)$ contains maximum energy among the other Slepian sequences and it is orthogonal to the previous sequence $\mu_0(t)$ and so on. So the slepian sequences are exactly band-limited and are the set of orthogonal sequences.

IV. ANALYSIS OF CHANNEL ESTIMATION

We assumed the following:

(A1): The input data sequence $\{b_n(i,k)\}$ is equi-powered with zero mean and variance E_b.

(A2): The additive noise $\{v^{(m)}(i,t)\}$ is white, uncorrelated with $\{b_n(i,k)\}$ having $E[v^{(m)}(i,t)] = \sigma_v^2$.

(A3): The LTV channel coefficients $h_{n,l}^{(m)}$ are complex Gaussian variables, and statistically independent.

By (A1)-(A3), the weighted average channel estimator mean square error (MSE) is given by

$$MSE_{n,l}^{(m)} = \{\left\| h_{n,l}^{(m)} - \hat{h}_{n,l}^{(m)} \right\|^2\} \tag{14}$$

We obtain variance of the weighted average estimation $\hat{h}_{n,l,q}^{(m)}$ as

$$\rho_{n,l,q}^{(m)} = \frac{(Q+1)E_b}{BE_p l^2} \sum_{i=1}^{l} \sum_{g=1}^{N} \sum_{k=1}^{L-1} \left\| h_{g,k}^{(m)}(i) \right\|^2 \tag{15}$$

The additive noise $v^{(m)}(i)$, i=1,…,I variance can be written as

$$E\left[\left\| v^{(m)} \right\|^2\right] = \frac{\sigma_v^2(Q+1)}{BlE_p} = \frac{\sigma_v^2(Q+1)}{\Omega E_p} \tag{16}$$

From equations (15) and (16), the variance of weighted average estimation can be written as

$$MSE_{n,l}^{(m)} = \frac{(Q+1)E_b}{\Omega l E_p} \sum_{i=1}^{l} \sum_{g=1}^{L-1} \sum_{k=1}^{L-1} \left| h_{g,k}^{(m)}(i) \right|^2 + \frac{\sigma_v^2(Q+1)}{\Omega E_p} \tag{17}$$

The estimation variances of the weighted average estimator were significantly reduced, so we can be considered the weighted average operation as an effective method for LTV channel estimation.

V. PERFORMANCE SIMULATION

Parameter	Specification	Value
Transmitter antennas	N	2
Receiver antennas	M	2
Symbol rate	f_s	10^7/second
Symbol size	B	512
Mobility speed	V	162km/h
Carrier frequency	f_c	2GHz
Frame size	Ω	131072
OFDM symbols	I	256
Channel delay	L	10
Basis expansion order	Q	10
Cyclic prefix length	$\overline{L}$	20

The transmitted data $s_n(i,k)$ is 8-PSK signals with symbol rate f_s. Before transmission, the transmitted data are coded by 1/2 convolutional coding and block interleaving over one OFDM symbol. We assume the channel delay L=10 and the coefficients of channel $h_n^{(m)}(t)$ are produced as zero mean random process, Gaussian, low-pass and are correlated with time according to the Jakes' mode,

$$r_n(\tau) = \mu_n^2 J_0(2\pi f_n \tau), \quad n = 1, \ldots, 2. \tag{18}$$

From the above equation Doppler frequency f_n is related with the nth user. The multi-path intensity profile is chosen to be $\varphi(l) = e^{l/10}, l = 0, \ldots, L-1$.

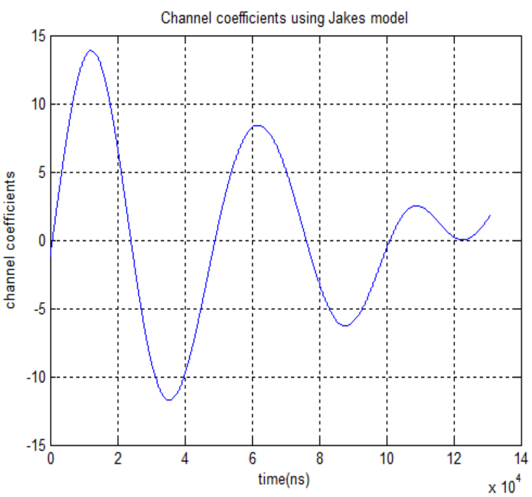


Fig.1. One tap coefficient of the LTV channel over the frame length Ω=131072

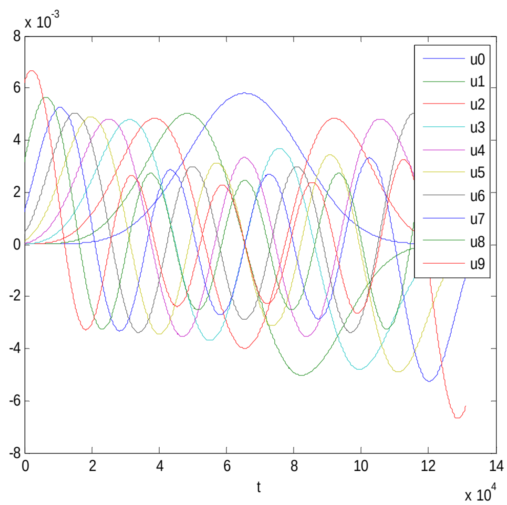


Fig.2. Slepian sequences $\mu_q(t)$ over frame length Ω=131072.

The Doppler spectra are $\psi(l) = \pi\sqrt{f_d^2 - f}$ for $f \le f_d$ where Doppler frequency of the different users as f_d, else, $\psi(l) = 0$. For avoiding the inter symbol interference (ISI), we choose the cyclic prefix length is 20. The noise is additive, white random processes, Gaussian and mean is zero.

The simulations can be run with the Doppler frequency f_n = 300Hz. The LTV channel is modeled with the frame size as $\Omega = B' \times 256 = (B + CP_{leng}) \times 256 = 131072$. Here we consider 256 OFDM symbols for every frame. During the frame, variation of the channel is $f_n\Omega / f_s$ = 4.1. In order to estimate the MIMO/OFDM channels, the superimposed pilots are designed according to (11) with the pilot power = 0.05.

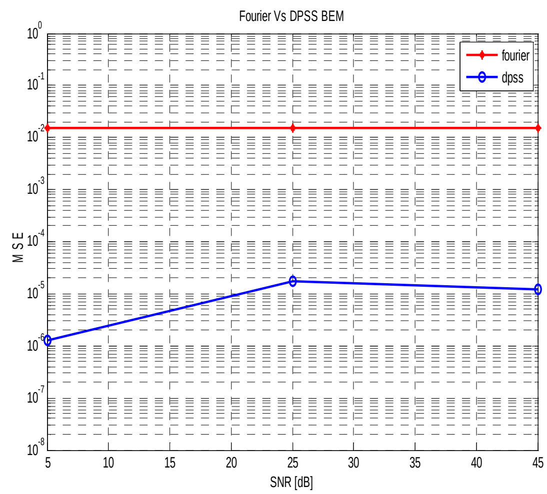


Fig.3. Comparison between Fourier BEM and DPSS BEM models for f_n=300Hz, Ω=13.62ms.

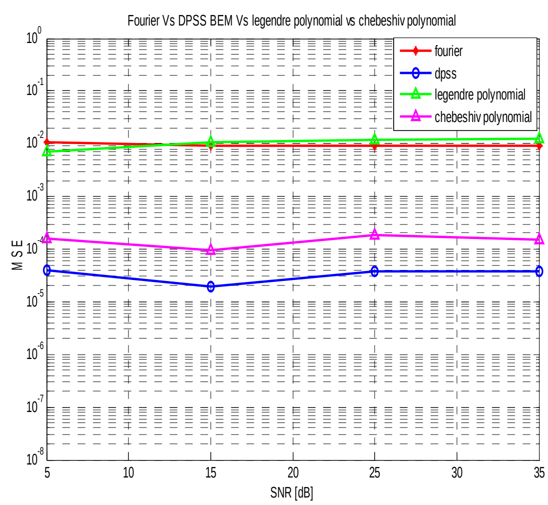


Fig.4 Mean square error versus Signal to noise ratio for Fourier BEM, DPSS BEM, Legendre and chebishev polynomial.

Fig.1 depicts the LTV channel coefficient estimation over the frame Ω=131072. It is clearly observed that although the channel coefficient is accurately estimated during the centre part of the frame. The Slepian sequences $\mu_q(t)$ for q={1,2,….10} are shown in Fig. 2. To measure the performance of the channel estimation, the mean square errors are used.

$$MSE_n^{(m)} = \sum_{i=1}^{\Omega/B'} MSE_n^{(m)}(i)$$

$$= \frac{B'}{\Omega} \sum_{i=1}^{\Omega/B'} E\left\{ \frac{\left| \sum_{t=0}^{B-1} \sum_{l=0}^{L-1} h_{n,l}^{(m)}(i,t) - \sum_{q=0}^{Q} \hat{h}_{n,l,q}^{(m)} e^{\frac{-j2\pi(q-Q/2)t}{\Omega}} \right|^2}{BL \left| h_{n,l}^{(m)}(i,t) \right|^2} \right\} \quad (19)$$

Where $\hat{h}_{n,l,q}^{(m)}$ is the estimation of channel coefficient and $MSE_n^{(m)}(i)$ denotes the mean square error of the i^{th} OFDM symbol. Fig.3 shows that the comparison of Fourier basis expansion and Slepian basis expansion. The simulations above reveal that the channel estimation performance can be improved using Discrete Prolate Spheroidal Sequences BEM model compared with the Fourier bases expansion model, along with the increment of training power.

VI. CONCLUSION

In this paper modeling of linear time varying channels for MIMO/OFDM system using different basis expansion models were compared. The channel coefficients were modeled using truncated DFB, and slepian sequences. We also present a performance analysis of Fourier BEM, DPSS-BEM, Legendre and chebishev polynomial. Simulation results shows that the mean square error performance is more for DPSS model compared with the Fourier basis expansion model.

REFERENCES

[1] G.B. Giannakis, and C.Tepedelenlioglu,"Basis expansion models and diversity techniques for blind identification and equalization of Time-varying channels," proc.IEEE, vol.86, pp. 1969-1986, oct.1998.

[2] I.Barhumi,G.Leus, and M.Moonen,"Optimal training design for MIMO OFDM systems in mobile wireless channels,"IEEE Trans. Signal process., vol.51, no.6, pp. 1615-1624, June 2003.

[3] T.Zemen and C.F. Mecklenbrauker,"Time-variant channel estimation using discrete prolate spheroidal sequences,"IEEE Trans signal process. vol.53, no.9,pp. 3597-3607,sep 2005.

[4] Q.Yang and K.S Kwak," Superimpose Pilot-aided channel estimation for mobile OFDM," Electron. Lett., vol.42, no.12, pp.722-724, June 2006.

[5] Xianhua Dai, Han Zhang, and Dong Li, "Linearly Time-Variant Channel Estimation for MIMO/OFDM systems using superimposed training" IEEE Trans. Commun., vol.58, No.2, pp.681-693, Feb 2010.

[6] L. Song and J.K. Tugnait, "On designing time-multiplexed pilots for doubly-selective channel estimation using descrete prolate spheroidal basis expansion models," in Proc. IEEE ICASSP Conf., Honolulu, HI, Apr.2007,pp. III -433- III -436.

[7] W.Hou and B.Chen," ICI cancellation for OFDM Communication systems in Time-varying multipath fading channels," IEEE Trns.Wireless Commun., vol.4, no.5,pp. 2100-2110, sep.2005.

[8] J. G. Proakis, Digital Communications, 4th ed. New York: McGraw-Hill, 2001.

[9] H. Kim and J, K. Tugnait, "Doubly-selective MIMO channel estimation using exponential basis models and subblock tracking , "in Proc. 42nd Annu. Conf. Inf.Sci. Syst., Mar. 19-21, 2008, pp. 1258-1261.

[10] N.Chen and G.T.Zhou,"Superimposed training for OFDM: a peak-to-average power ratio analysis," IEEE Trans. Signal Process., vol.54, no.6, pp.2277-2287, June 2006.

[11] T.Keller and L.Hanzo, "Adaptive multi carrier modulation: a convenient frame work for time-frequncy processing in wireless communications," proc. IEEE, vol.88, no.5, pp. 611-640, May 2000.

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 3, March 2016

Investigating the Distributed Load Balancing Approach for OTIS-Star Topology

Ahmad M. Awwad, Jehad Al-Sadi

Abstract— *This research effort investigates and proposes an efficient method for load balancing problem for the OTIS-Star topology. The proposed method is named OTIS-Star Electronic-Optical-Electronic Exchange Method; OSEOEM; which utilizes the electronic and optical technologies facilitated by the OTIS-Star topology. This method is based on the previous FOFEM algorithm for OTIS-Cube networks. A complete investigation of the OSEOEM is introduced in this paper including a description of the algorithm and the stages of performing Load Balancing, A comprehensive analytical and theoretical study to prove the efficiency of this method, and statistical outcomes based on common used performance measures has been also presented. The outcome of this investigation proves the efficiency of the proposed OSEOEM method.*

Keywords— *Electronic Interconnection Networks, Optical Networks, Load balancing, Parallel Algorithms, OTIS-Star Network.*

I. INTRODUCTION

THE Star graph was proposed by Akers and *et al* as one of the most promising graph due to its attractive topologies. The Star graph has excellent topological properties when we compare it with networks of similar sizes [1, 2]. The Star graph shown to have many attractive properties over many networks such as the well-known cube network, including: smaller diameter, smaller degree, and smaller average diameter [1]. The Star graph proved to have a hierarchical structure which will enable it to construct larger network size from multi smaller ones [1, 2].

With the new advances of technology, new era of Optical networks has been appeared in literature. Many previous researches addressed the emerging of the Optical technology with the traditional electronic interconnection topologies. OTIS-Star was one of the proposed networks in this era due to its attractive properties and features [3].

Although some algorithms proposed for the OTIS-Star graph such as routing algorithms and distributed fault-tolerant routing algorithm [3, 4], still there is a shortage of efforts to solve the problem of load balancing algorithm that utilizes the OTIS-Star networks attractive topologies.

To our knowledge there is inadequate results proposed in literature about implementing and proposing efficient algorithms for load balancing on OTIS-Star topology. In this paper we try to fill this gap through proposing and embedding the OSEOEM algorithm on the OTIS-Star graph which is based on the FOFEM algorithm which was shown to be efficient on OTIS-Cube networks [5]. The main mechanism of this algorithm is to redistribute the load equally as possible among the processors of the network [6]. Efficient implementation of the OSEOEM algorithm on the OTIS-Star network will make it more suitable network for real life application in connection to load balancing problem. The rest of the paper is organized as follows: In section 2 we present the necessary basic notations and definitions, in section 3 we introduce some of the related work on load balancing, in section 4 we present and discuss the implementation of the OSEOEM algorithm on the OTIS-Star graph, also we present an example of **OSEOEM** on OTIS-3-Star network, in section 5 we present and discuss some analytical study for the proposed load balancing algorithm. Section 6 conducts a performance study on the statistical and numerical results issued. Finally section 7 concludes this paper.

II. DEFINITIONS AND BASIC TOPOLOGICAL PROPERTIES

A quit big number of interconnection networks for HSPC appeared by many researchers in the last decade [7, 9], such networks including Star and the OTIS-Star interconnection networks. The OTIS-Star topology [3] is a well-known example of these new appeared networks; it has been presented as new promising network to Star network [1]. From the time of it has been presented, the OTIS-Star network attracted a lot of research studies, few topological properties of this attracted network have been presented in the literature such as: its basic topological properties [1], optimal parallel path [8], distributed algorithms of the load balancing [6] and embedding of other topologies [10]. Authors of [3] have utilized the attractive properties of the factor graph and its superiority over the other similar network showing an efficient load balancing and fault-tolerant parallel paths algorithms. The studied properties include lower degree, a smaller both diameter and a smaller average diameter [1, 2].

The topological properties and hierarchy of the proposed OTIS-Star network enables an efficient algorithm design step for the proposed graph OTIS-Star. This graph may be seen as $n! \times n!$ as it has been introduced by authors in [8], where the symbol (!) stands for the factorial of n, also in the naming of the algorithm each node has a unique permutation on $\langle n \rangle = \{1,\ldots,n\}$ of Star graph.

Limited researches have been introduced in the literature related to building efficient algorithms for the OTIS-Star interconnection network including algorithms of broadcasting, routing, and load balancing [5, 11, 12, 13, 14]. This paper attempts to solve the shortage of limited research efforts by presenting an efficient algorithm to redistribute the load size among all the nodes of different groups within the OTIS-Star topology, this redistribution will allow balancing of load size among all processors of the network as equally as possible.

The topological properties of the OTIS-Star topology along with those of the Star network are discussed below. These topological properties have been discussed using the theoretical framework for analyzing topological properties of the OTIS-Networks in general which was proposed by Al-Ayyoub and Day [2], beside other related research work [11].

Throughout this paper, the symbol g represents the group address and the symbol p represents the processor address. An optical link which connects two different groups is formed as $(\langle g, p\rangle, \langle p, g\rangle)$ which means an OTIS connection.

Multiplying any factor topology by itself will result in an OTIS-network of the factor topology. The vertex set in the new resulted OTIS-Factor topology is achieved by performing Cartesian product on the set of vertices of the factor network. On the other hand, the edge set of the achieved OTIS-Factor consists of edges from the factor network and transpose edges which connect edges between different groups. A definition of the OTIS-n-Star graph is presented in the next paragraph.

The OTIS-Star network has N groups; each group has N number of nodes, which leads to a total of N^2 nodes in the whole network. Each node is addressed by the notation of $\langle x, y\rangle$, such that $0 \leq x, y < N$ where x is the group address and y is the processor address. Note that any two nodes in the same group are connected by the factor interconnection link; while any two nodes of different groups are connected through an optical link which achieved by swapping group address with processor address of the two nodes.

***Definition* 1:** The OTIS-n-Star Graph, which is denoted by OTIS-n-Star has $n! \times n!$ nodes each addressed as a unique representation of the permutation $\langle n\rangle = \{1\ldots n, 1\ldots n\}$. This address consists of two parts where the first part represents the group address and the second part represents the factor address of the node within the group. For any two nodes of the OTIS-factor network to be connected, their corresponding representation must differ exactly in the first position and any other position of their address representation.

***Definition* 2:** In the undirected Star graph, the factor network can be represented as n-Star = (V_0, E_0). And the OTIS-n-Star network represented by (V, E) which it is an undirected graph obtained from n-Star as follows $V = \{\langle x, y\rangle \mid x, y \in V_0$ such that $x \neq y\}$ and $E = \{(\langle x, y\rangle, \langle x, z\rangle) \mid$ if $(y, z) \in E_0\} \cup \{(\langle x, y\rangle, \langle y, x\rangle) \mid x, y \in V_0$ such that $x \neq y\}$.

The OTIS methodology implements n-Star edges by electronic links and implements transpose edges through the different groups. In this research the "*electronic move*" and the "*OTIS move*" will be used to point to data transmission utilizing both electronic and optical technologies, respectively.

Figure 1 shows the OTIS-3-Star graph with 6 groups; 3! = 6; each has 6 nodes.

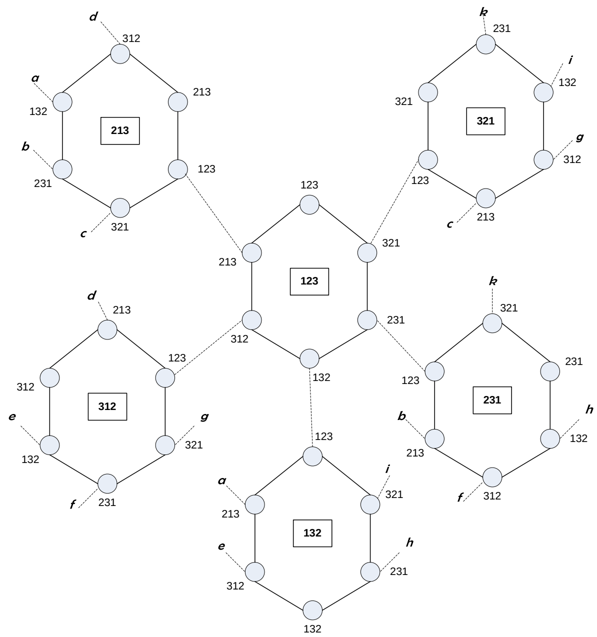


Fig. 1: The OTIS-3-Star Graph

III. Background and Related Work

The OTIS-n-Star graph has showed an attractive properties which already been presented and published by many researchers in the literature. Based on the above, we have been motivated to extend our research efforts on the OTIS-n-Star network, in specific to investigate the load balancing problem. We have concentrated on the load balancing algorithm since there is a huge gap of research on these kinds of problems for the OTIS topologies. The load balancing problem has been studied and proposed on different kinds of infrastructure including the electronic networks [5] and OTIS networks [6].

The load balancing is well-known and it is one of the important kinds of problems which were studied on different types of interconnection graph topologies. Ranka, Won, and Sahni [15] were the first to present a study of this problem on the hypercube topology; they investigated and proposed a new algorithm called the Dimension Exchange Method (DEM). The proposed algorithm was built on the concept of calculating the average load of nodes which are connected directly as neighbors. As an example the of network with dimension n, the set of neighbors on the n^{th} dimension their load balancing will be exchanged to redistribute the load among the nodes to achieve evenly load balancing as possible. In DEM algorithm, the nodes that have extra load will redistribute this load to its adjacent neighbor's nodes. The main attractive point of DEM algorithm is that all the nodes will redistribute any task to its adjacent neighbors to come up with equally load balancing between these nodes. Researchers in [15] have concluded that the worst efficiency of the DEM algorithm in redistributing the load balancing was $log_2 n$ for the cube network.

Furthermore, the authors of [16, 17] have introduced another algorithm for load balancing on the OTIS interconnection networks, this algorithm is called Diffusion and Dimension Exchange (DED-X). The algorithm works by grouping the load balancing for any task to three different stages. The achieved outcomes on the OTIS network proved that the efficiency of load balance is approximately redistributed equally among the nodes of the network. Also the same authors have presented a simulation work were the achieved results of the simulation have shown a major improvement and enhancement of the load balancing issue on OTIS-networks [16, 17]. The authors of [18] have introduced a method called DED-X for load balancing on homogeneous OTIS networks. The DED-X algorithm is based on a structure called Generalized Diffusion-Exchange-Diffusion Method, this scheme allowed the load balancing on different OTIS networks [18].

Moreover the SCDEM algorithm which was introduced in [6] based on the Clustered Dimension Exchange Method CDEM for load balancing for OTIS on Hypercube factor network [6]. The efficiency of CDEM for load balancing on OTIS-Hypercube calculated as $O(\text{Sqrt}(n)*M \log_2 n)$ where M is the maximum load assigned to each processor in a n-processor Hypercube. On the other hand the number of communication steps that is required by CDEM is $3log_2\, n$ [6].

Day and Al-Ayyoub have proposed a new load balancing algorithm for OTIS-Cube networks [5], they have shown that the usability of the new proposed load balancing algorithm is more effective compared to the introduced load balancing algorithm DED-X [18].

IV. The OSEOEM Algorithm

The objective of this section is to propose a new load balancing algorithm for the OTIS-n-Star networks called OSEOEM. The proposed algorithm is based on the well-known algorithm FOFEM which was introduced in [5].

OSEOEM algorithm aims to achieve equally load balancing for the OTIS-n-Star network through redistributing the assigned tasks of the nodes among the different nodes of different groups of the factor networks. In this algorithm the number of exchanges done between different nodes in the OSEOEM is $2(n!)+1$, where n is the degree of factor network; n-Star. The algorithm for load balancing problem on OTIS- *3*-Star network is presented in Figure 2.

The proposed algorithm consists of three phases as follows:

- Phase 1: This phase aims to redistribute the load size among all nodes within each factor group of the network as evenly as possible. This phase is performed in n stages. The 1st stage of this phase is performed via redistributing and balances the load size of all adjacent neighbor nodes such that there is a difference in the permutation address of any two adjacent nodes in the 1st and 2nd position of their addresses.

The 2nd stage of this phase is achieved one more time through the redistributing the load size of any two direct neighbour nodes that their corresponding permutations differ exactly in the first and 3rd position.

The n^{th} stage is also achieved through redistributing the load size among the 1st position and nth position, where n is the last symbol of the node address of the n-Star factor network.

Generally phase 1 will continue in redistributing the load size of all the neighbor nodes that their permutations differ in the first and the i^{th} position, where $2\leq i \leq n$. This phase will be repeated two times until we achieve an optimal redistribution of load size among all nodes within each group of the n-Star network.

- Phase 2: In this phase, an exchange to redistribute the load size of any two adjacent that are connected optically is conducted. At the end of this phase the load size will be redistributed almost equally between all the groups, at the end of this stage we will have different node load size in the group itself which at the end will leads to the need of conducting the third phase.
- Phase 3: This phase will repeat the same stages of phase 1 as shown in Figure 2. The aim of this stage is to achieve an approximately even load size among the nodes of the factor network in each group. Noting that phase 2 has changed this even distribution done in phase number 1 in order to redistribute the load size in the whole groups of the network. This justifies the need for phase 3 to order the redistribution in the factor network once again.

Note that n-1 is the number of neighbors' of any processor in the factor network S_n:

A detailed description of the OSEOEM phases in shown in Figure 2:

1. *for* $m = 2$; $m \leq n$; m++ // *Start of phase 1*
2. *for* all n-1neighbour nodes; p_i and p_j which they differ in 1st and m position of S_n do in parallel
3. Give-and-take p_i and p_j total load sizes of the two nodes
4. *TheAverageLoad* $p_{i,j}$ = Floor ((*Load* p_i +*Load* p_i)/2)
5. *if* (*Totalload* p_i >= excess *AverageLoad* $p_{i,j}$)
6. *Send* excess load p_i to the neighbour node p_i
7. Load p_i = Load p_i – extra load
8. Load p_j = Load p_j + extra load
9. *else*
10. Receive extra load from neighbour p_j
11. Load p_i = Load p_i + extra load
12. Load p_j = Load p_j – extra load
13. Repeat steps 3 to 12 one more time // end of phase 1
14. *for* all adjacent nodes via an optical link, exchange the loads of the nodes // phase 2
15. Redistribute the weight within each group by repeating steps 1 – 13 // phase 3

Fig. 2: The OSEOEM load balancing Algorithm

The three phases of OSEOEM algorithm are performed in order to redistribute and balance the load size among all nodes of the network. A detailed description of the three phases of the algorithm is stated as follows:

- Phase 1: OSEOEM algorithm performs a load balancing and redistribution of load size between the nodes of S_n via performing steps 2 to 12 of the algorithm in parallel. The first step aims to redistribute the load size of all the nodes in each group which they differ in 1st position and 2nd position in the nodes' addresses in the same group; the factor graph S_n. Then this process is repeated continuously to redistribute the load size of all the nodes in each group which they differ in 1st position and n^{th} position to redistribute the load size among all nodes in each factor group. Phase 1 is repeated again in order to achieve an approximate equal load balancing among all nodes in any group, this repetition will guarantee the redistribution of nodes at distance n-1 in their 1st and n^{th} address position of their addresses; border nodes.

- Phase 2: This phase is conducted in parallel and it aims to redistribute the load size among all the groups of the graph by exchanging loads of any two nodes that are connected optically.
- Phase 3: This phase repeats all steps of phase 1 to redistribute load size of all nodes within each factor graph one more time since phase 2 disordered the loads within any group in order to evenly exchange the load across the groups of the whole network.

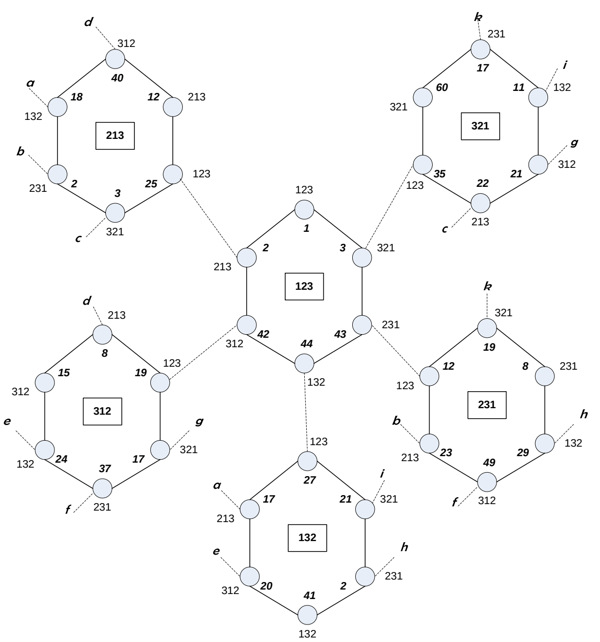


Fig. 3: OTIS-3-Star network – load balancing - initial state

The following example will illustrate OSEOEM algorithm.

Example: -This example demonstrates the load balancing size on the OTIS-3-Star where the factor network is S_3.

Firstly, OTIS-3-Star network is shown in Fig. 3, where the factor network S_3 consists of 6 nodes, the whole network has 3! groups of S_3. The load size of each node is presented in bold and italic which is assigned next to each node represents the initial load. Any node is connected to neighbouring nodes within a group via electronic links; furthermore this node is connected to a third node in another group via an optical link.

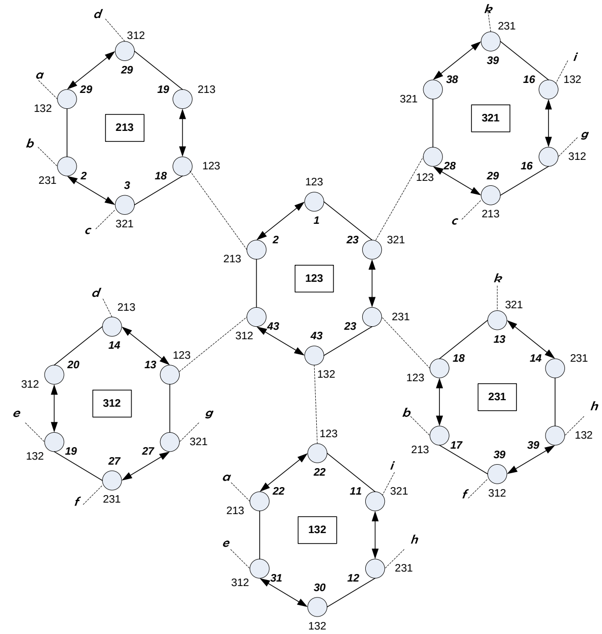


Fig. 4: OTIS-3-Star network – load balancing phase 1 step 1

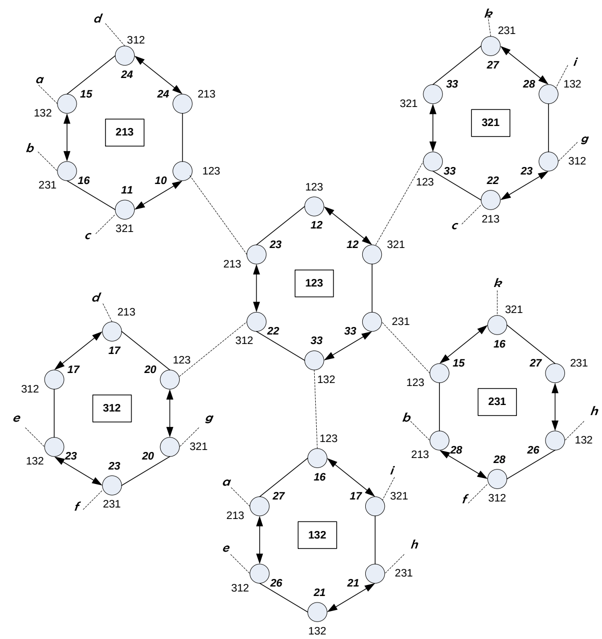


Fig. 5: OTIS-3-Star network – load balancing phase 1 step 2

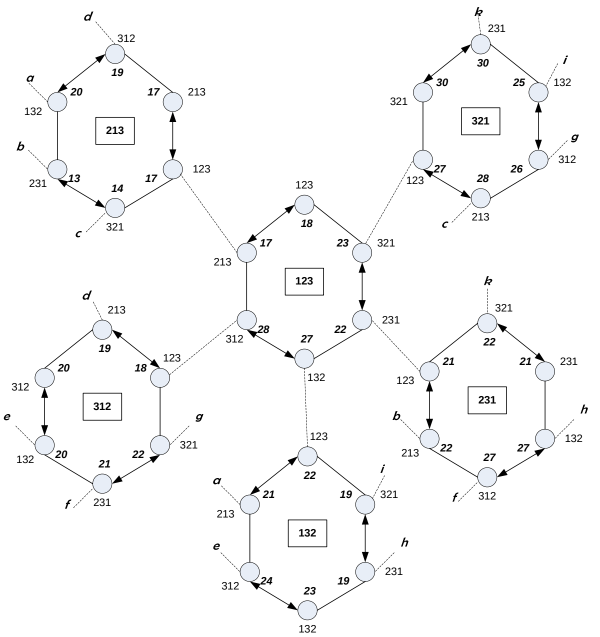


Fig. 6: OTIS-3-Star network – load balancing phase 1 step 3

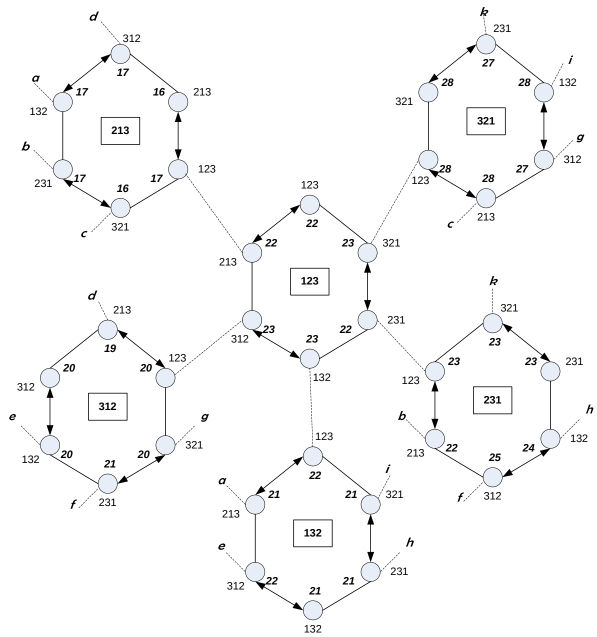


Fig. 8: OTIS-3-Star network – load balancing phase 1 step 5

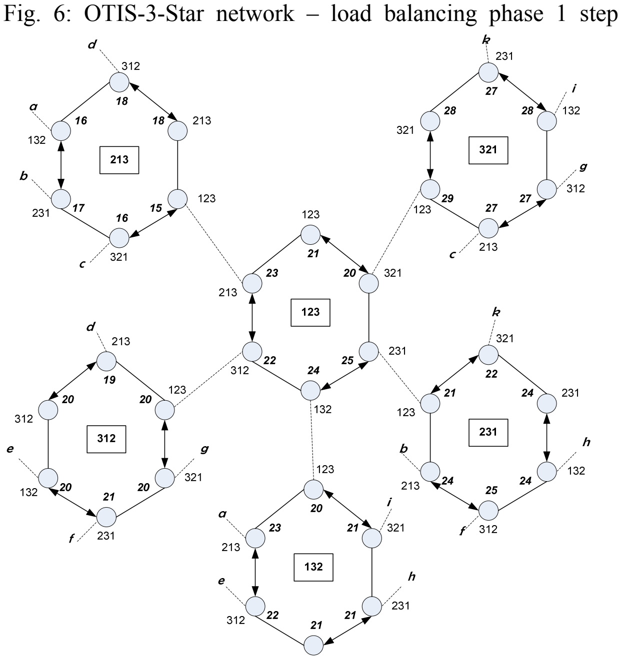


Fig. 7: OTIS-3-Star network – load balancing phase 1 step 4

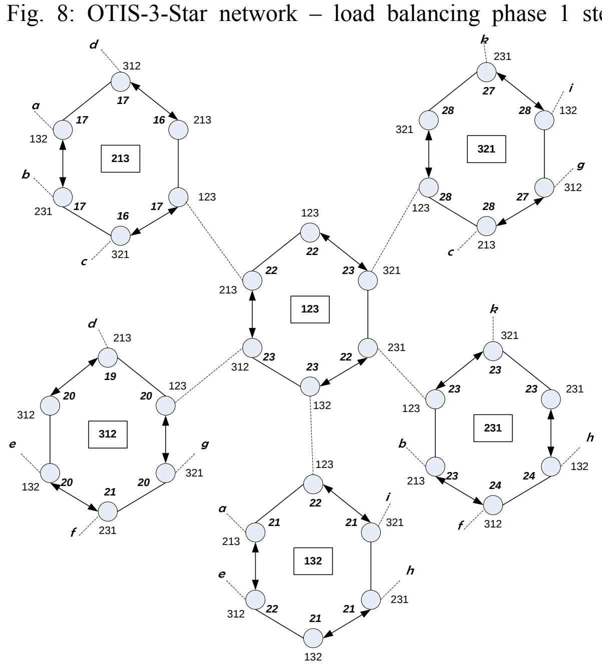


Fig. 9: OTIS-3-Star network – load balancing phase 1 step 6

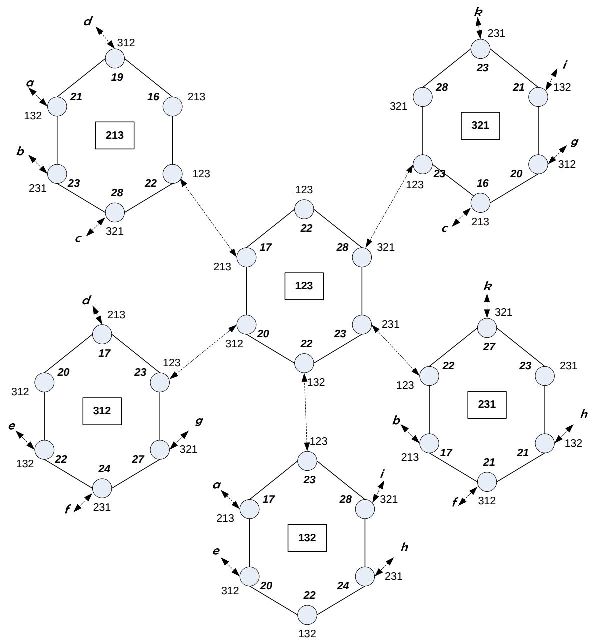


Fig. 10: OTIS-3-Star network – load balancing phase 2

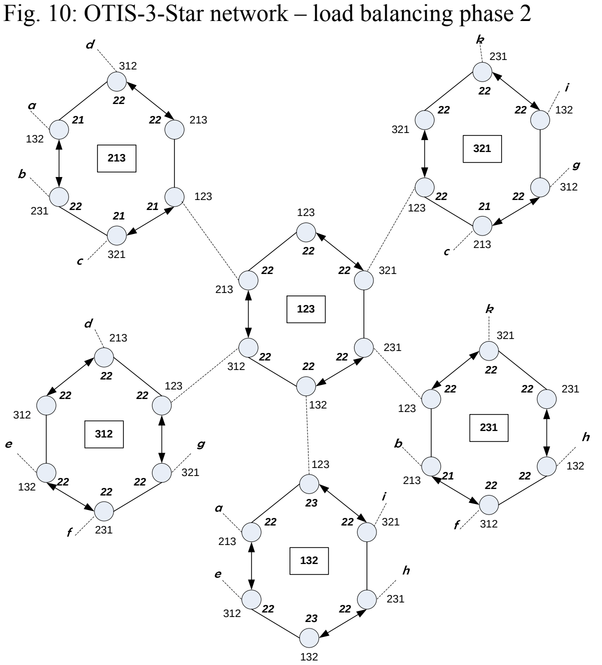


Fig. 11: OTIS-3-Star network – load balancing- end of phase 3

The OSEOEM algorithm starts implementing phase 1 via performing the steps 2-12 of the algorithm. Figures 4 and 5 illustrate the first step of this phase. Every two nodes which they are differ in 1st position and any other position will redistribute their load size almost equally. The first step is repeated two more times to enhance and produce more accurate distribution, the first repetition of phase 1 is shown in Figures 6 and 7, while Figures 8 and 9 illustrate the second repetition of this phase.

In Phase 2, every two nodes that are connected to each other directly via an optical link exchange their load size in order to balance the load sizes across the groups; the implementation of this phase is shown in Figure 10.

Finally, after performing phase 3, the final outcome of OSEOEM algorithm is shown in Figure 11. This figure shows that all nodes will have almost the same load size which proves the efficiency of the proposed algorithm.

The final result of OSEOEM algorithm is shown to be efficient and optimal. The final distribution is achieved in $2 \times n! + 1$ communication steps where n is the degree of the OTIS-n-Star network.

V. Analytical Study

In this section we will present an analytical study to prove that the proposed algorithm is efficient in terms of low number of permutations, execution steps, and latency time

Theorem 1

For the OTIS-n-Star network to reach mostly an even load balancing among different nodes at least $2 \times n!+1$ permutations will be required.

Proof:-

In the factor network, the number of local nodes is $n!$ in every group of the n-Star factor groups that formulate the whole OTIS-n-Star topology. To redistribute the load size inside any factor Star group, a maximum of $\frac{n!}{2}$ parallel exchanges are needed to exchange the load between any two nodes via an optimal distance path.

An additional $\frac{n!}{2}$ of parallel exchanges are needed to make sure an accurate redistribution of load size is done among all the factor nodes including the nodes at diameter distance, this leads to guarantee that almost the same load size distributed among all nodes in the factor n-Star group in at most $n!$ parallel exchanges.

By the end of the $n!$ parallel exchanges, all factor n-Star groups will have almost an equally distribution among all nodes locally within each group.

Furthermore, another one parallel exchange is needed to exchange the loads between every two groups of the OTIS-n-Star topology via an optical link that connects these groups, the main purpose of this parallel exchange is to guarantee that every group will have almost same load size of other groups, Also it is important to know that this step will disorder the load size of all nodes within the local groups..

Finally, to reorder the load size locally within each group, we need another $n!$ exchanges in parallel to redistribute the load size among all nodes at each group locally, this means that the $n!$ parallel exchanges are done before and after the on parallel optical exchange.

The total number of parallel exchanges of the above steps we be 2×n! + 1 which prove the theorem.

Theorem2:

In the OTIS-n-Star network, the OESOEM algorithm performs a total of $\frac{n!(n!-1)}{2}$ number of parallel optical exchanges that occurs simultaneously.

Proof:-

Since the OTIS-n-Star topology has n! groups, and every group is connected to the other n!-1 groups via only one optical link between any two groups, this means that there are a total number of $n!\,(n!-1)$ optical connections.

But it is known that every optical link connects every two groups bi-directionally, this require to divide the total number of connection by 2, so we conclude that the total number of parallel optical exchanges will be $\frac{n!(n!-1)}{2}$ that is done simultaneously.

Theorem3:

The time required to execute OSEOEM algorithm on the OTIS-n-Star is $2 \times n! \times t_e + t_o$ when the latency time is discarded where:

t_e is the maximum time required to exchange the load via an electronic link between any two nodes locally. And t_o is the maximum time required to exchange the load via an optical link between any two groups,

Proof:-

By referring to theorem 1, $(2 \times n!)$ electronic steps are needed to perform the factor exchanges of OSEOEM, each exchange will be performed at a maximum time of t_e. Then the total time required for this stage is $t_e \times 2 \times n!$, additionally, the one optical exchanges occurs as one parallel step, the time needed to perform this optical step is t_o; this leads to an overall time $t_e \times 2 \times n! + t_o$ which is required.

Lemma 1:

$l_e \times 2 \times n! + l_o$ is the latency time to accomplish the load balancing OSEOEM algorithm on the OTIS-n-Star topology, where: l_e is the needed latency time to transmit the load between any two nodes in the factor topology. Furthermore l_o is the maximum time wanted to transmit the load from between any two nodes via optical technology.

Lemma 2:

To execute the OSEOEM load balancing in the OTIS-n-Star topology, a total time of $(t_e + l_e) \times 2 \times n! + t_o + l_o$ is needed.

Furthermore to minimize the total time needed to execute the OSEOEM load balancing in the OTIS-n-Star network, we can minimize the total number of electronic communication steps from $n!$ to $\frac{(n!)}{2} + 1$ for each of the two phases needed in the OSEOEM algorithm (phase 1 and phase 3). We argue that any exchange of data between any two nodes in the Star topology can be achieved in $\frac{(n!)}{2}$ steps as discussed in theorem 1.

Also we will need $\frac{(n!)}{2}$ more steps to guarantee an accurate equal distribution the load among nodes in the factor topology. We can make the algorithm more flexible by doing only one exchange step of $\frac{(n!)}{2} + 1$ exchanges instead of n! exchanges at each phase, this will distribute the load size close to the equality but with minor differences of load sizes, the extra 1 step in the above equation is to make sure that the load sizes is redistributed among the diameter distance nodes, this is a trade of between flexibility and accuracy. Also we argue that after doing the one optical exchange in phase 2, there will be another $\frac{(n!)}{2}$+1 exchanges of load balancing among all nodes in the factor topology at phase 3. The total number of parallel exchanges for the whole process will be minimized from 2×n!+1 to n!+3.

This will lead to an approximate acceptable distribution of load sizes among the nodes, experimental results showed that the upper-bound variation of load sizes after minimizing the exchanges steps will not exceed 2 units of the load size variation if we implement the 2×n!+1 exchanges in the proposed OSEOEM algorithm. By shortening this amount of exchanges, this will lead to minimizing the total time to perform this algorithm (the actual communication time and also the latency time), this means we will enhance the performance of algorithm but at the expense of accuracy.

Figure 12 shows the final load distribution after minimizing the number of exchanges for pervious example.

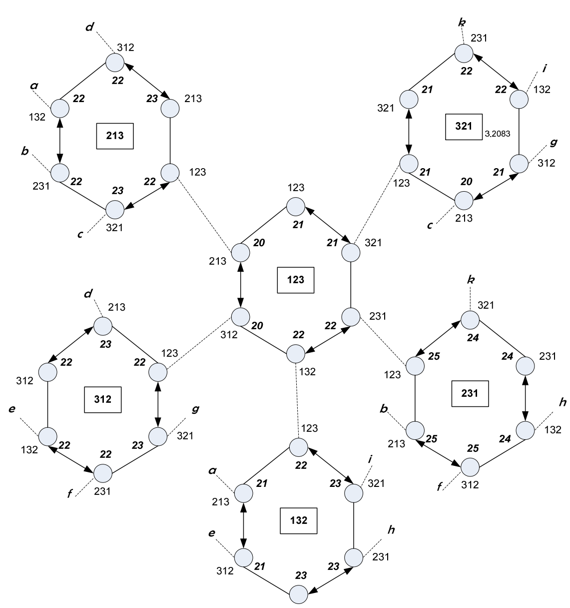


Fig. 12: OTIS-3-Star network – load balancing- end of phase 3 with minimized exchanges

It is obvious from figure 12, that the load balancing varies from one node to another one with a maximum variation of 2, if we compare it with the original algorithm, but we save the time to perform this load balancing by reducing the number of needed exchanges. The next section will present and experimental comparison for different network sizes to show the effect of this reduction of exchanges on the approximate time to perform the load balancing.

VI. Performance Study

In this section we will present statistical and numerical results of a set of experiments to show the impact of the above theorems on the algorithm and the topology itself.

Table 1 shows the number of exchanges in the OTIS-*n*-Star by implementing the OSEOEM load balancing algorithm, these exchanges does not represent the total number of exchanges since some exchanges occurs in parallel, so it represent the sequential exchanges as it occurs in a sequence of steps. The table shows that the number of exchanges is very small compared with the size of the OTIS-*n*-Star topology. The last column presents the percentages between number of nodes of the network and the number of exchanges required by the OSEOEM algorithm. When the network size gets larger the percentage of exchanges gets smaller, this fact proves the efficiency of OSEOEM algorithm.

Table 1. OSEOEM total number of exchanges for OTIS-*n*-Star

N	# of nodes *n*-Star	# of nodes OTIS-*n*-Star	total # of exchanges	percentage size/exchanges
	n!	$(n!)^2$	2 *n*!+1	
3	6	36	13	36.1111111%
4	24	576	49	8.5069444%
5	120	14400	241	1.6736111%
6	720	518400	1441	0.2779707%
7	5040	25401600	10081	0.0396865%
8	40320	1625702400	80641	0.0049604%
9	362880	1.31682E+11	725761	0.0005511%
10	3628800	1.31682E+13	7257601	0.0000551%
11	39916800	1.59335E+15	79833601	0.0000050%
12	479001600	2.29443E+17	958003201	0.0000004%

In Theorem 3, we presented the time required to complete the three phases of OSEOEM algorithm which contain both optical and electronic time needed. Table 2 present the time needed to complete the algorithm by ignoring the latency time, we assume that the maximum time required to transmit the load size between any two nodes via an electronic link is 250 Mb/s [19, 20]. We also assume that the maximum time required to transmit the load from a source node to a destination node via an optical link is 2.5 Gb/s [21]. The second column in the table is fixed since there is only one optical move, same observation from table 1 is applied here, the total time required to perform OSEOEM gets smaller when the network size gets larger if we compare it with the number of nodes at each network size.

Table 2. OSEOEM total required time for OTIS-*n*-Star

n	Electronic required time	Optical required time	Total required time
	250Mb/s	2.5Gb/s	
3	3000	2.5	5560
4	12000	2.5	14560
5	60000	2.5	62560
6	360000	2.5	362560
7	2520000	2.5	2522560
8	20160000	2.5	20162560
9	181440000	2.5	181442560
10	1814400000	2.5	1814402560
11	19958400000	2.5	19958402560
12	2.39501E+11	2.5	2.39501E+11

To calculate the total time required to perform OSEOEM algorithm, we need to calculate the latency time; as we already introduced the latency time in Lemma 1; table 3 presents the total latency time which include the optical and electronic latency time. We assume that the maximum latency time required to transmit the load from a source node to a destination node via an electronic link is 4.76 Micro seconds. Also we assume that the maximum time required to transmit the load from a source node to a destination node via an optical link is 0.07 Micro seconds [22].

Table 3. OSEOEM total latency time for OTIS-*n*-Star

n	*Electronic latency time*	*Optical latency time*	*Total latency time*
	4.76 Micro Sec	0.07 Micro Sec	Micro Sec
3	0.000057	0.00000007	0.0000572
4	0.000228	0.00000007	0.0002286
5	0.001142	0.00000007	0.0011425
6	0.006854	0.00000007	0.0068545
7	0.047981	0.00000007	0.0479809
8	0.383846	0.00000007	0.3838465
9	3.454618	0.00000007	3.4546177
10	34.546176	0.00000007	34.5461761
11	380.007936	0.00000007	380.0079361
12	4560.095232	0.00000007	4560.0952321

The total time required to perform OSEOEM is the summation of total execution time and total latency time for each network size.

VII. Conclusion

This paper presented an efficient algorithm for load balancing of nodes in OTIS-*n*-Star network. The proposed algorithm is called OSEOEM which is based on the known FOFEM algorithm presented for OTIS-Cube network. The proposed OSEOEM algorithm performs load balancing among the all nodes of OTIS-*n*-Star network by redistributing of load sizes almost equally across all nodes. The algorithm redistributes load balancing among all nodes in $2(n!)+1$ communication steps which is considered to be efficient.

Furthermore, this research paper presents an analytical and performance study on the proposed algorithm which includes the communication steps, percentage of exchanges, execution time, and latency time to prove the OSEOEM is an efficient mathematically.

References

[1] S. B. Akers, D. Harel and B. Krishnamurthy, "The Star Graph: An Attractive Alternative to the n-Cube" *Proc. Intl. Conf. Parallel Processing*, 1987, pp. 393-400.

[2] K. Day and A. Tripathi, "A Comparative Study of Topological Properties of Hypercubes and Star Graphs", IEEE Trans. Parallel & Distributed Systems, vol. 5.

[3] J. AL-Sadi, A. M. Awwad, B. F. AlBdaiwi, Efficient Routing Algorithm on OTIS-Star Network, Proceedings of the IASTED International Conference on Advances in Computer Science and Technology, Virgin Islands USA, November 22-24, 2004, ACTA Press, pp. 157 – 162.

[4] Khaled Day and Abdel-Elah Al-Ayyoub, "Node-ranking schemes for the Star networks", Journal of parallel and Distributed Computing, Vol. 63 issue 3, March 2003, pp 239-250.

[5] Jehad Al-Sadi, "Implementing FEFOM load balancing algorithm on the enhanced OTIS-n-Cube topology", The Second International Conference on Advances in Electronic Devices and Circuits - EDC2013, Kuala Lumpur, Malaysia, May, 2013.

[6] B.A. Mahafzah and B.A. Jaradat, "The Load Balancing problem in OTIS-Hypercube Interconnection Network", J. of Supercomputing (2008) 46, 276-297.

[7] S. B. Akers, and B. Krishnamurthy, "A Group Theoretic Model for Symmetric Interconnection Networks," *Proc. Intl. Conf. Parallel Proc.*, 1986, pp. 216-223.

[8] K. Day and A. Al-Ayyoub, "The Cross Product of Interconnection Networks", *IEEE Trans. Parallel and Distributed Systems*, vol. 8, no. 2, Feb. 1997, pp. 109-118.

[9] A. Al-Ayyoub and K. Day, "A Comparative Study of Cartesian Product Networks", *Proc. of the Intl. Conf. on Parallel and Distributed Processing: Techniques and Applications*, vol. I, August 9-11, 1996, Sunnyvale, CA, USA, pp. 387-390.

[10] I. Jung and J. Chang, "Embedding Complete Binary Trees in Star Graphs," *Journal of the Korea Information Science Society*, vol. 21, no. 2, 1994, pp. 407-415.

[11] Berthome, P., A. Ferreira, and S. Perennes, "Optimal Information Dissemination in Star and Panckae Networks," *IEEE Trans. Parallel and Distributed Systems,* vol. 7, no. 12, Aug. 1996, pp. 1292-1300.

[12] P. Fragopoulou and S. Akl, "A Parallel Algorithm for Computing Fourier Transforms on the Star Graph," *IEEE Trans. Parallel & Distributed Systems*, vol. 5, no. 5, 1994, pp. 525-31.

[13] Mendia V. and D. Sarkar, "Optimal Broadcasting on the Star Graph," *IEEE Trans. Parallel and Distributed Systems,* Vo;. 3, No. 4, 1992, pp. 389-396.

[14] N. Imani et al, "Perfect load balancing on Star interconnection network", J. of supercomputers, Volume 41 Issue 3, September 2007. pp. 269 – 286.

[15] Ranka, Y. Won, S. Sahni, "Programming a Hypercube Multicomputer", IEEE Software, 5 (5): 69 – 77, 1998.

[16] Zhao C, Xiao W, Qin Y (2007), "Hybrid diffusion schemes for load balancing on OTIS networks", In: ICA3PP, pp 421–432

[17] G. Marsden, P. Marchand, P. Harvey, and S. Esener, "Optical Transpose Interconnection System Architecture," *Optics Letters*, 18(13), 1993, pp. 1083-1085.

[18] Qin Y, Xiao W, Zhao C (2007), "GDED-X schemes for load balancing on heterogeneous OTIS networks", In: ICA3PP, pp 482–492.

[19] R Kay Pragmatic Network Latency Engineering Fundamental Facts and Analysis, CPacket Networks, URL: http://www. cpacket. com , 2009 pp 1-31.

[20] Kibar O, Marchand P, Esener S., "High speed CMOS switch designs for free- space optoelectronic MINs.", IEEE Transactions on Very Large Scale Integration (VLSI) Systems 1998; 6(3):372–386.

[21] Esener S, Marchand P., "Present and future needs of free-space optical interconnects". I*n 15 IPDPS 2000 Workshop on Parallel and Distributed Processing*, 2000; 1104–1109.

[22] Farrington,N. et al, "A Multiport Microsecond Optical Circuit Switch for Data Center Networking", *IEEE PHOTONICS TECHNOLOGY LETTERS*, VOL. 25, NO. 16, 2013, pp 1589-1592.

American Sign Language Pattern Recognition Based on Dynamic Bayesian Network

Habes Alkhraisat, Saqer Alshrah

Abstract—**Sign languages are usually developed among deaf communities, which include friends and families of deaf people or people with hearing impairment. American Sign Language (ASL) is the primary language used by the American Deaf Community. It is not simply a signed representation of English, but rather, a rich natural language with a unique structure, vocabulary, and grammar. In this paper, we propose a method for American Sign Language alphabet, and number gestures interpretation in a continuous video stream using a dynamic Bayesian network. The experimental result, using RWTH-BOSTON-104 data set, shows a recognition rate upwards of 99.09%.**

***Index Terms*— American Sign Language (ASL), Dynamic Bayesian Network, Hand Tracking, Feature extraction.**

I. Introduction

TIGN languages are developed among deaf communities and people with hearing impairment. The sign languages development depends on the community and the region. Consequently, they are vary from region to region and also they have their own grammar, e.g. there exists American Sign Language and German sign language. However, there is no relation between sign language in a particular region to the spoken language.

Sign language includes different components of visual actions of the signer made by using the hands, the face, and the torso, to convey the meaning. The information in sign languages is expressed by hand/arm gestures including position, orientation, con-figuration and movement of the hands .These gestures are named manual components of the signing and convey the meaning of the sign.

Manual components has two categories: glosses and classifiers. Glosses are the signs for language word. Classifiers are designated handshapes and/or rule-grounded body pantomime to represent nouns and verbs. The classifier provides additional information about nouns and verbs such as location, kind of action, size, shape and manner. ASL has many classifier handshapes to represent specific categories or class of objects. Classifiers are used to express movement, location, and appearance of a person or a subject. Signers use classifiers to express a sign language word by explaining where it is located and how it moves or what it looks like by its appearance.

During signing a sentence, the hand(s) need to move from the ending location of one sign to starting position of the next sign. In addition, the hand configuration changes from ending hand configuration of one sign to the starting of the next. This movement is called movement epenthesis, and although it happens frequently between the signs, it does not belong to the components of sign language.

A sign language recognition system is a considered as the key of a communication between deaf or hard hearing people and hearing people. It includes a hardware for data acquisition to extract the features of the signings, and a decision-making system to recognize the sign language.

To signing extract features, most researches in ASL required special data acquisition tools like data and colored gloves, location sensors, or wearable cameras. In contrast, the proposed system in this article is designed to recognize sign language words and sentences directly from the frames captured by standard cameras using simple appearance-based features and geometric features of the signers' dominant hands. Therefore, this system can be used rather easily in practical environments. The main goal behind this work is to build a robust hand-based sign language recognition system.

II. American Sign Language

The Sign language is a visual language, words are produced by moving the hands combined with facial expressions and postures of the body to communicate and receive information. ASL contains pronunciation, vocabulary, and grammar and syntax [1] [2]. The Sign language is a shift from "listening for language" to "looking for language".

The vocabulary of ASL consists of signs, which are the analog of words in spoken language. The handshape, location, movement, palm orientation, and non-manual signals are the essential elements of a signs [3]. The most apparent and complex parameter of a sign is the hand shape [4]. The hand shapes in ASL are composed of letter, number, and classifier hand shapes [5] [6]. The Hand shapes are used as the "index" in dictionaries that facilitate the lookup of an unknown sign. Another essential part of ASL is fingerspelling, which used for spelling proper nouns, acronyms, and technical terms and would not be possible without hand shapes.

Hand shapes are particular configurations of the hand; a relatively small set (40) generates the majority of signs in ASL [7]. Comprehension of a sign depends on recognizing the hand shape. For example, the ASL signs for "year" and "world" have the same pattern of movement but differing hand shapes.

I. Fingerspelling

Fingerspelling is a method of spelling words with hand shapes and hand movements. Fingerspelling is used in sign language to spell out names of people and places for which there is not a sign. Fingerspelling can also be used to spell words for signs that the signer does not know the sign for, or to clarify a sign that is not known by the person reading the signer. Fingerspelling signs are often also incorporated into other signs.

Two types of manual alphabet are used around the world. The one handed alphabet is the most common and is used in sign languages such as American Sign Language (ASL) (figure 1 and 2), Irish Sign Language (ISL), and many of the European sign languages. The two handed manual alphabet is used by British Sign Language (BSL), Australian Sign Language (AUSLAN), New Zealand Sign Language (NZSL), among others.

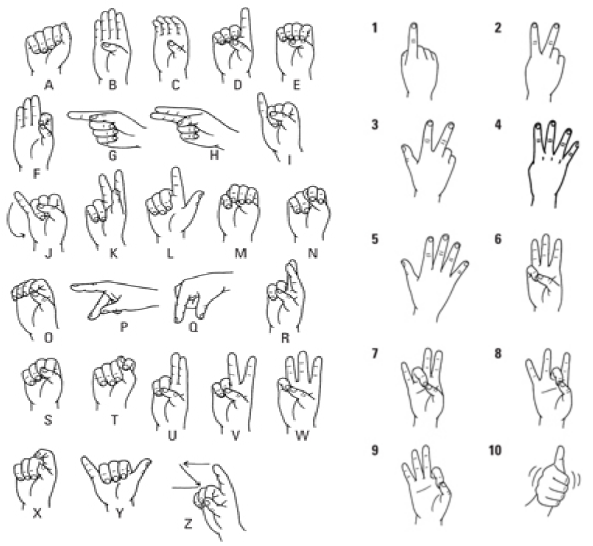


Fig. 1. ASL Hand Shapes.

II. ASL FACIAL EXPRESSIONS

In addition to the gestural communications of hand, signs, and fingerspelling the facial expressions are a key component of ASL. The meaning of the sign is made up from the facial expressions together with a sign. For example, if you sign the word "quiet," and add an exaggerated or intense facial expression, you are telling your audience to be "very quiet." The same principle also works when making "interesting" into "very interesting," or "funny" into "very funny." Facial expressions are called non-manual markers that influence the meaning of signs. Non-manual markers include facial expressions, head tilt, head nod, headshake, shoulder raising, mouth morphemes, and other non-signed signals

Facial expression is a key component of grammar that involves far more than merely emotional disposition or level of intensity. The gestural complexity of ASL necessitates the use of the head and facial features as an intrinsic part of the language. Facial expressions are used in conjunction with word signs and fingerspelling to communicate specific vocabulary, questions, intensity, and subtleties of meaning [2].

Fig. 2. Phrases in American Sign Language.

III. HANDS TRACKING

To extract manual features, the hands are tracked in each image sequence. Location of hand and track it in space-time the key success of ASL recognition and influences on the performance of the recognition system. In our proposed method, we use the Real-time tracking of multiple skin-colored objects with a possibly moving camera [8] [9]. In [8] [9], the camera captures an image, the skin-colored blobs are detected and maintains a set of object hypotheses that have been tracked up to this instance in time. The detected blobs and object hypotheses are then associated in time, to assign a new unique label to each new object that enters the camera's field of view for the first time, and to propagate in time the labels of already detected objects.

For Skin, color detection involves (a) estimation of the probability of a pixel being skin-colored, (b) hysteresis thresholding on the derived probabilities map, (c) connected components labeling to yield skin-colored blobs and, (d) computation of statistical information for each blob. Skin color detection adopts a Bayesian approach, involving an iterative training phase and an adaptive detection phase [8].

The blob of a hand and face is A described using Gaussian distribution and the mean represents the location of the hand or face centroid (Figure 3). Each blob b_j, $1 \leq j \leq M$, corresponds to a set of connected skin-colored image points. For linear prediction, the optical flow, which measures the motion explicitly across frames, has been used. The optical flow measures the motion explicitly across frames so that it can still succeed in tracking.

Once the optical flow between the previous frame and the current frame has been computed for a blob, Gaussian mean of the blob in the current frame is predicted using the average of the optical flow vectors:

$$v = \frac{1}{N} \sum_{i=1}^{N} f(i) \qquad (1)$$

where $\boldsymbol{f}(i) = [f_x(i), f_y(i)]^T$ denotes the $i-th$ flow vector and N the number of flow vectors associated with the blob.

Fig. 3. Hand tracking

IV. Feature Extraction

The feature extraction of hand motion is an important procedure for ASL recognition. The accuracy of automated ASL recognition is highly influenced by the nature of extracted features and the employed matching process. The motion can be described by the trajectory in space over time, which in turn is represented by a sequence of positions or equivalently motions vectors x_t, $t = 1, \cdots$. The location x_t at time t is estimated by the mean of the Gaussian fitting the blob. Each pair of successive hand locations defines a local motion vector. After that, the whole motion trajectory is represent by a sequence of motion vectors each of which is encoded by a direction code (Figure 3(a)). The central code '0' denotes 'no motion'. Given a video, we extract two chain codes one for each hand.

To remove the ambiguities between gestures incurred by representing the motion using only the chain code, two more features is used: the two hands relative position (Figure 4(b)) and the position of the each hand relative to the face (Figure 4(c)). The code '0' implies that two hands, a hand, and a face are overlapping.

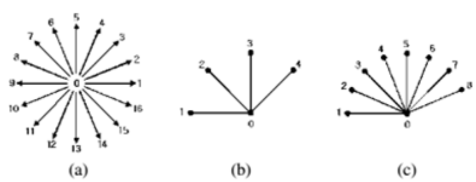


Fig. 4. Features (1) 17 directions code for hand motions, (b) hand-hand positional relation, and (c) face-hand positional relation.

V. Dynamic Bayesian Network (DBN)

A Bayesian Network (BN) graphically represents the relations between a set of random variables, it represents a joint probability distribution of a set of random variables with a possible mutual causal relationship. It consists of two major parts: a directed acyclic graph and a set of conditional probability distributions. The directed acyclic graph consists of nodes, and edges. The nodes represent the random variables, and the edges between pairs of nodes representing the causal relationship of these nodes, and a conditional probability distribution in each of the nodes [10].

Dynamic Bayesian Networks (DBNs) generalize HMMs by allowing the state space to be represented in factored form, instead of as a single discrete random variable. DBNs generalize KFMs by allowing arbitrary probability distributions, not just unimodal linear-Gaussian. By using DBNs, it is possible to represent, and learn, complex models of sequential data, which are closer to "reality".

DBN model probability distributions over semi-infinite collections of random variables $Z1, Z2, Z3, \cdots$. To represent the input, hidden and output variables of a state-space model, the variables are partitioned into $Z_t = (U_t, X_t, Y_t)$.

A DBN is defined to be a pair, $(B1, B \rightarrow)$, where $B1$ is a BN which defines the prior $P(Z1)$, and $B \rightarrow$ is a two-slice temporal Bayes net (2TBN) which defines $P(Zt|Zt-1)$ by means of a directed acyclic graph as follows:

$$P(Z_t|Z_{t-1}) = \prod_{i=1}^{N} P(Z_t^i|Pa(Z_t^i)) \quad (2)$$

where Z_i^t is the ith node at time t, which could be a component of X_t, Y_t or U_t, and $Pa(Z_i^t)$ are the parents of Z_i^t in the graph. The nodes in the first slice of a 2TBN do not have any parameters associated with them, but each node in the second slice of the 2TBN has an associated conditional probability distribution (CPD), which defines $P(Z_t^i|Pa(Z_t^i)$ for all $t > 1$.

For example, consider four random variables $w, x, t,$ $and\ z$. From basic probability theory, we know that we can factor the joint probability as a product of conditional probabilities:

$$P(w, x, y, z) \\ = P(w)P(x|w)P(Y|w, x)P(z|w, x, y) \quad (3)$$

Each variable is represented by a node in the network. A directed arc is drawn from node X to node Y if Y is conditioned on X in the factorization of the joint distribution. For example, Figure 5 represents the Bayesian network for the following factorization:

$$P(W, X, Y, Z) = P(W)P(X)P(Y|W)P(Z|X, Y) \quad (4)$$

Hidden Markov models fall in a subclass of Bayesian networks known as dynamic Bayesian networks, which are simply Bayesian networks for modeling time series data. In time series modeling, the assumption that an event can cause another event in the future, but not vice-versa, simplifies the design of the Bayesian network: directed arcs should flow forward in time. Assigning a time index t to each variable, one of the simplest causal models for a sequence of data $\{Y_1, \ldots, Y_t\}$ is the first-order Markov model, in which each variable is directly influenced only by the previous variable:

$$P(Y_{1:T}) = P(Y_1)P(Y_2|Y_1) \cdots P(Y_T|Y_{T-I}) \quad (5)$$

Although the HMM [11] is a very useful tool for modeling variabilities, its power is limited to a very simple state space with a single discrete hidden variable. The coupled HMM is an HMM variant tailored to represent the interaction of two independent processes [12]. It is essentially two HMMs coupled between the state variables across the HMMs. Although useful for modeling simple interacting processes, this model does not have room for common hidden variables, which are believed to be shared between two variables, and is often hard to extend due to the exponential computation as the number of coupled processes increases. It is also hard to add new information into these models because of the rigid structure.

The dynamic Bayesian network (DBN) [13] is a general generalized framework of HMM and Bayesian network (BN). With an appropriate design, it can make up for the weaknesses of the HMM by factorizing the hidden variable into a set of random sub-variables.

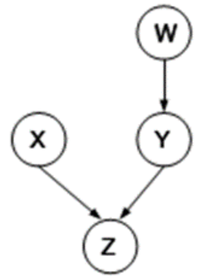


Fig. 5. A direct acyclic graph (DAC)

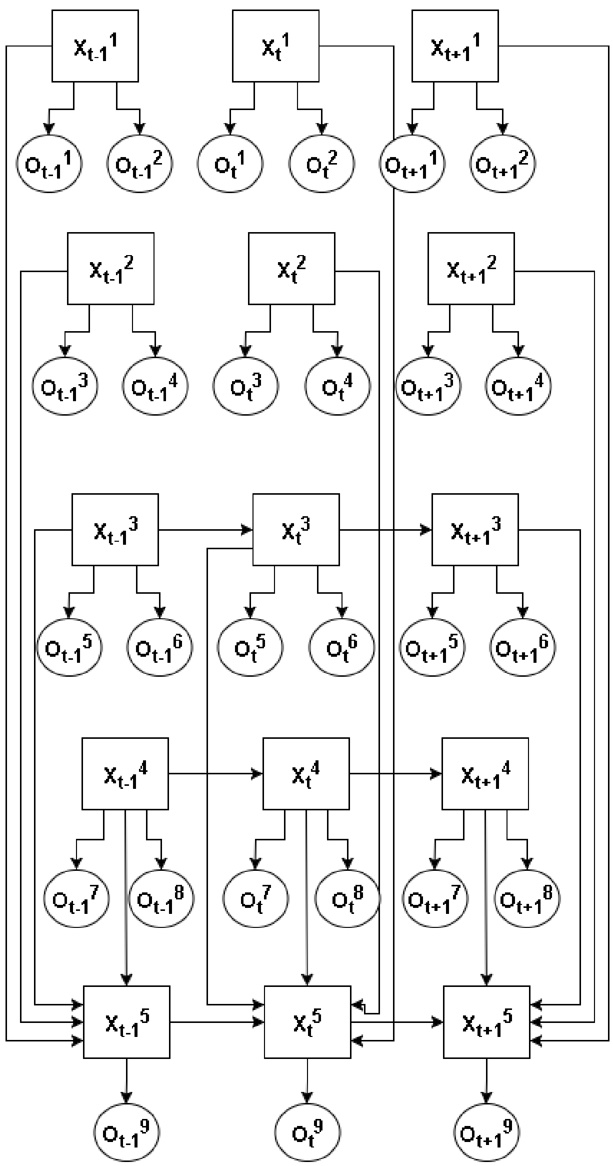


Fig. 6. The dynamic Bayesian network model for ASL recognition

VI. PROPOSED MODEL ARCHITECTURE

We are proposing a new design of DBN, which has five hidden variables and nine observable variables. The two hidden variables X_1 and X_2 model the motion of the left and the right hand respectively, and each variable is associated with two observations of the features of the corresponding hand's motion and the position relative to the face. The third hidden variable X_3 has been introduced to resolve the ambiguity between similar sign. The hidden variables X_4 model the facial expression and with two observations of the features of the corresponding heads motion and the facial expression. The hidden variable X_5 has been introduced to resolve the ambiguity between similar sings.

Figure 6 illustrates the propose ASL recognition model, where the hidden variables in square nodes and observable variables in circle nodes.

VII. INFERENCE

The DBNs allows computing the joint probability of a subset of variables very efficiently. The goal of inference in a DBN is to compute the margin probability of hidden variable X_i given an observation sequence.

The joint probability of variables in a BN can be factored into a product of local conditional probabilities one for each variable through conditional independencies or d-separation. The full joint probability for the DBN in Figure 6 is computed by multiplying five factored probabilities as follows:

$$P(X_{1:T}^{1:3}, O_{1:T}^{1:9}) = P(O_{1:T}^{1:9} | X_{1:T}^{1:N}) P(X_{1:T}^{1:9}) \tag{6}$$

where $X_{1:T}^{1:5} = \begin{bmatrix} X_1^1 \\ \vdots \\ X_1^5 \end{bmatrix} \cdots \begin{bmatrix} X_T^1 \\ \vdots \\ X_T^5 \end{bmatrix}$

and $O_{1:T}^{1:9} = \begin{bmatrix} O_1^1 \\ \vdots \\ O_1^9 \end{bmatrix} \cdots \begin{bmatrix} O_T^1 \\ O\vdots \\ O_T^9 \end{bmatrix}$

The nodes in each time slice are sufficient to d-separate the past from the future. Therefore, if the values of all nodes in time-slice t are given, the nodes in the next time slice at $t + 1$ are independent of all those preceding t. Therefor only, the two time-slice Bayesian network (2TBN) at each time $t = 1$ is considered. In 2TBN, the value of the nodes which have outgoing arcs to the next time-slice is sufficient to d-separate the past and the future. Then the number of nodes reduce 1.5TBN. The interface refers to nodes d-separateing the past from the future [13]. Then a junction tree is built where the interface must be included the same subset of nodes in a graph such that there exists a link between all pairs of nodes in the subset, those subset of nodes is called 'clique'. Once a junction tree is built, the junction tree algorithm (JTA) is applied [14]. The JTA follows a message-passing protocol. The message updating from clique C_i to clique C_j can be computed as follows:

$$F = \sum_{C_i / S_{ij}} \mathrm{flod}(C_i) \tag{7}$$

$$f(C_i) = \mathrm{flod}(C_i) \times \frac{f(S_{ij})}{\mathrm{flod}(S_{ij})} \tag{8}$$

where, S_{ij} is the separator which includes intersecting nodes of C_i and C_j, $C_i \setminus S_{ij}$ a set difference, and f a potential function.

VIII. Method description

The proposed method is designed to recognize American Sign Language words and sentences using simple skin color features and geometric features of the signers' dominant hand and face, which are extracted directly from the frames captured by standard cameras (figure 7). The proposed method for American Sign Language recognition operates as follows: for feature extraction, the head and the hands of signer have to be found. To extract features that describe manual components of a sign, the hand has to be tracked in each image sequence. At each time instance, the camera acquires an image on which skin-colored blobs (i.e. connected sets of skin-colored pixels) are detected. The method also maintains a set of object hypotheses that have been tracked up to this instance in time. The detected blobs, together with the object hypotheses are then associated in time.

The goal of this association is (a) to assign a new, unique label to each new object that enters the camera's field of view for the first time, and (b) to propagate in time the labels of already detected objects.

Three hidden variables and five observable variables DBNs has been employed to recognize continuous American Sign Language sentences. The DBN recorded the recognition rate of 99.09%.

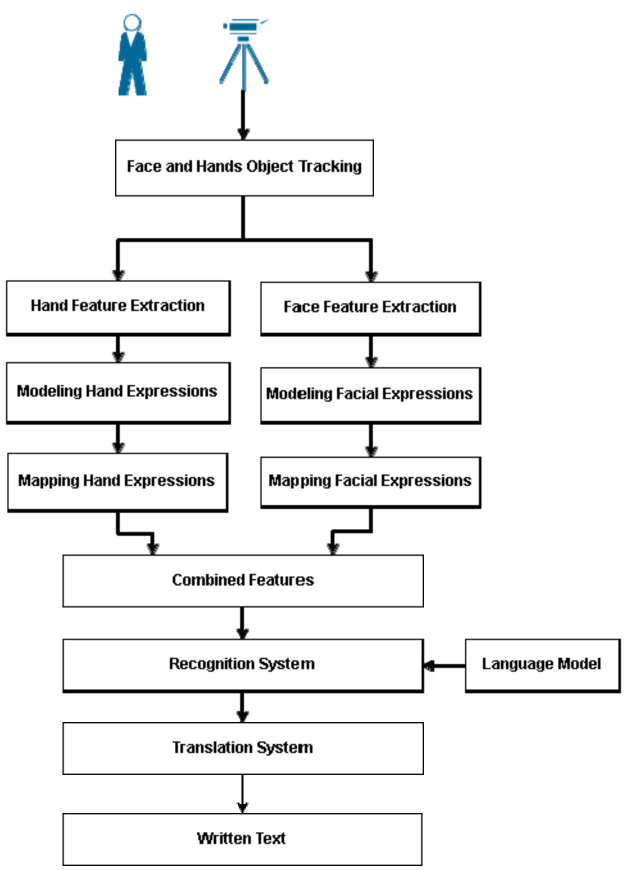


Fig. 7. The dynamic Bayesian network model for ASL recognition

IX. Experimental Results

The experiments for American Sign Language recognition have been performed on the publicly available RWTH-BOSTON-104 database. In the RWTH-BOSTON databases, there are three signers: one male and two female signers. All of the signers are dressed differently and the brightness of their clothes is different. It consists of 201 annotated video streams of ASL sentences and these video streams can be used for sign language recognition. On the average, these sentences consist of five words out of a vocabulary of 104 unique words.

Table 1
Corpus statistics for the RWTH-Boston-104 database

	Training Set		Evaluation Set
	Training	Development	
Number of sentences	131	30	40
Number of running words	568	142	178
Vocabulary size	102	64	65
Number of singletons	37	38	9
Number of OOV words	-	0	1

Table 1 shows the corpus statistics for BOSTON-104 database, which include number of sentences, running words, unique words, singletons, and out-of-vocabulary (OOV) words in the each part. Singletons are the words occurring only once in the set. The out-of-vocabulary words are the words, which occur only in the evaluation set, i.e. there is no visual model for them in training set and they cannot be therefore recognized correctly in the evaluation process. Based on the experiments with RWTH-BOSTON-104 database, the DBN recorded the recognition rate 99.09%.

X. Conclusion

In this work, we proposed an automatic American Sign Language (ASL) recognition system based on Dynamic Bayesian Network. The proposed method have three hidden variables, which together take five observations: chain codes of each hand's motion, relative position between the face and each hand, and relative position of two hands. We tested the DBN-based system performance with a RWTH-BOSTON-104 database. The DBN model showed the recognition rate of 99.09%.

References

[1] C. Valli, and C. Luca, 1993. Linguistics of American Sign Language, Galluadet University Press.

[2] S. Baker-Shenk, and D. Cockely, 1980. American Sign Language, A Teacher's Resource Text on Grammar and Culture, Gallaudet University Press.

[3] S. Liddell, and R. Johnson, 1989. American Sign Language: The Phonological Base, Sign Language Studies, 64, 195-277.

[4] D. Brentari, a Prosadic Model of Sign Language Phonology. Cambridge, MA: MIT Press.

[5] C. Valli, and C. Lucas, 1995. Linguistics of American Sign Language: An introduction. Washington, D.C.: Gallaudet University Press.

[6] A. Grieve, and B. Smith, 1998. Sign Synthesis and Sign Phonology, Proc. The First Annual High Desert Linguistics Society, Albuquerque:HDLS.

[7] R. Tennant, and M. Brown, 1980. The American Sign Language Handshape Diction-ary, Clerc Books, Washington D.C.

[8] A. Argyros and M. Lourakis, 2004. Real-time tracking of multiple skin-colored objects with a possibly moving camera, Proc. of European Conference on Computer Vision, pp. 368–379, 3.

[9] H. Suk, B. Son, and S. Lee, 2000. Recognizing Hand Gestures using Dynamic Bayesian Network. 8th IEEE International Conference on Automatic Face & Gesture Recognition, pp. 1-6.
[10] Murphy K. 1998. A Brief Introduction to Graphical Models and Bayesian Networks
[11] L. Rabiner, 1989. A tutorial on hidden markov models and selected applications in speech recognition. Proc. of the IEEE, pp. 257–285, 77.
[12] M. Brand, N. Oliver, and A. Pentland, 1997. Coupled hidden Markov models for complex action recognition, Proc. Of IEEE Computer Society Conference on Computer Vision and Pat-tern Recognition, pp. 994–999.
[13] K. Murphy, 2002. Dynamic Bayesian Network: Representation, Inference and Learning, PhD dissertation, University of California, Berkeley.
[14] C. Huang and A. Darwiche, 1994. Inference in belief networks a procedural guide. International Journal of Approximate Reasoning, 15(3), 225–263.

Habes Alkhraisat was born in Amman, Jordan, in 1979. He received the B.S. in information technology from Al-Balqa Applied University in 2001, and M.S. degrees in computer science from the University of Jordan, in 2003 and the Ph.D. degree in System Analysis, Information Processing and Control from Saint Petersburg Electro Technical University (LETE), Saint Petersburg, Russia, in 2008.

Since 2009, he has been an Assistant Professor with the Computer Science department, Applied University. His research interest includes the system modeling, image processing, and biometric.

Saqer M. Al-Shra'h has received B.S. in computer information system from the Hashemite University in 2010, and M.S. degrees in computer information system from the University of Jordan. Currently he is Trainer and head of IT department in institute of public administration. His research interest includes Network Security and image processing,

ISSN 1947-5500

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 3, March 2016

Identification of Breast Cancer by Artificial Bee Colony Algorithm with Least Square Support Vector Machine

S. Mythili, Research Scholar,
PG& Research Department of Computer Application,
Hindusthan College of Arts & Science,
Coimbatore, India

Dr. A. V. Senthilkumar, Director,
PG& Research Department of Computer Application
Hindusthan College of Arts &Science
Coimbatore, India

***ABSTRACT*: procedure for the identification of several discriminant factors. A new method is proposed for identification of Breast Cancer in Peripheral Blood with microarray Datasets by introducing the Hybrid Artificial Bee Colony (ABC) algorithm with Least Squares Support Vector Machine (LS-SVM), namely as ABC-SVM. Breast cancer is identified by Circulating Tumor Cells in the Peripheral Blood. The mechanisms that implicate Circulating Tumor Cells (CTC) in metastatic disease is notably in Metastatic Breast Cancer (MBC), remain elusive. The proposed work is focused on the identification of tissues in Peripheral Blood that can indirectly reveal the presence of cancer cells. By selecting publicly available Breast Cancer tissues and Peripheral Blood microarray datasets, we follow two-step elimination.**

***Keywords: Breast Cancer (BC), Circulating Tumor Cells (CTC), Peripheral Blood (PB), Artificial Bee Colony (ABC), Least Squares Support Vector Machine (LS-SVM*).**

I.INTRODUCTION

Circulating cancer cells have been detected in a majority of epithelial cancers tissues which includes breast, prostate, lung cancer. Patients with metastatic lesions are most likely to have CTCs detected in their blood tissues [1]. Recent studies in this BC have risen interesting mechanistic. For example, CTC captured in xenograft prostate BC models have highlighted the importance of pathways with conferring resistance to apoptosis in these cells [2]. Studies of the effects of Epithelial–Mesenchymal Transition (EMT) in the generation of CTCs and distal metastases have suggested that this mesenchymal transformation may enhance the ability of cells to intravasate but may reduce their competence to initiate over metastases [3-4]. Most of the studies have identified bone marrow–derived hematopoietic progenitor cells that express VEGF receptor 1 (VEGFR1) and may form a premetastatic niche that precedes the arrival of tumor cells in the blood tissues [5]. Moreover, [6] have newly proposed a concept of tumor self-seeding, in which injected tagged human cancer cell lines may colonize an existing tumor deposit, with the newly recruited tumor cells conferring increased aggressiveness to the existing tumor.

In addition, Barbazan et al. report that the spread of cancer relates to the detachment of malignant cells into blood [7] and Obermayer et al. [8] demonstrate that CTCs can be detected in single-cell level through specific genes (six gene panel) in PB. Particular microarray studies on PB that isolates specific CTC cells report that CTCs carry characteristics from the primary cause [8], but also convey information regarding the secondary metastasis tumor [6]. Moreover, some specific alterations in cancer might be indicative of its ability to diffuse; such genes can indirectly predict the existence of CTCs without the need to detect and/or extract them [9].

II.RELATED WORK

Several innovative approaches have been developed to detect these types of rare tumor cells. Some of these cells make use of interesting physical or biological properties of epithelial cells. High densities of microscopic scanning approaches have been adapted to screen for CTCs [10]. Laser-scanning cytometry method is used for combining the fluorescent labeling and forward scatter to enhance identification of tumor cells which is deposited on a glass slide [11]. In general, these studies have concluded that the presence of detectable CTCs in the blood serves as an independent prognostic factor in patients with Breast Cancers. In patients with metastatic breast cancer, CTC counts above five CTCs per 7.5 ml of blood before the start of systemic therapy were associated with a shorter median progression-free survival and overall survival [12-13]. Additional studies extended these analyses to use molecular endpoints, such as HER2 staining, and to patients with invasive localized breast cancer receiving so-called neoadjuvant chemotherapy [14-15]. However, despite processing as much as 50 ml of blood, CTCs were detected in only half of the patients, with the number of HER2-positive cells ranging from one to eight CTCs per 50 ml. Thus, although promising, these approaches emphasize the critical need for increased sensitivity in CTC detection to enable clinical applications.

Lin et al [16] developed a portable filter-based micro device that is both a capture and analysis platform capable of multiplexed imaging and genetic analysis and has the potential to enable routine CTC analysis in the clinical setting for the effective management of cancer patients. This device is based on the size difference between CTCs and human blood cells and has been reported to achieve CTC capture on filter with approximately 90% recovery within 10 min. The same group has developed and validated a novel 3-dimensional microfiltration device that can enrich viable CTC from blood. The device provides a highly valuable tool for assessing and characterizing viable enriched circulating tumor cells in both research and clinical settings [17].

III.PROPOSED METHOD

Hypothesis supports those specific differences of cancer tissue and cancer blood are indicative of the ability of tumor to diffuse and, thus, can be used as factors for CTC estimation without direct detection. For this purpose, in this work proposed a hybrid ABC-SVM procedure applied on several publicly available DNA microarray datasets from different origins (tissue and blood). The first stage aims to extract gene signatures associated with pair wise differentiation between cell types and/or disease states. For instance, the comparison of cancer and control tissue provides information about the discriminative factors of the primary disease. Next, proposed a hybrid classification method for the detection of CTC, between cancer blood and control PB in association with the primary and secondary disease. Overall, we consider the hypothesis that this intersection, representing the common features of primary tumor and BC PB, is likely to reflect CTCs biology.

1.1. Gene Differentiation

In the initial stage of this work, we used the SAM method [18] with the siggenes package of R/Bio conductor. It also uses the false discovery rate (FDR) [19] as the criterion for determining the set of genes that exhibit differential expression and its critical value has been set to 0.01 for all comparisons. The use of FDR implies that the resulting gene sets that were found to have differentiating expression values do not have the same number; instead, the number of differentially expressed genes differs among comparisons.

1.2. Hybrid classification method for CTC identification

Support Vector Machine (SVM) is known as a powerful methodology for solving problems in nonlinear classification, function estimation and density estimation. In this work SVM has been introduced to the detection of CTC and classifies them into meta and non metastasis within the context of statistical learning theory. Least squares support vector machine (LS-SVM) is reformulations from standard SVM which lead to solving linear Karush–Kuhn–Tucker (KKT) systems. LS-SVM is closely related to regularization networks and Gaussian processes but additionally emphasizes and exploits primal–dual interpretations [20]. In LS-SVM function estimation, the standard framework is based on a primal–dual formulation. Given gene dataset samples for CTC identification with N dataset $\{x_i, y_i\}_{i=1}^{n}$, the goal is to estimate a model of the form

$$y(x) = w^T \varphi(x) + b + e_i \tag{1}$$

where $x \in R^n$, $y \in R$ and $\varphi(.) : R^n \to R^{n^h}$ is a mapping to a high dimensional gene dataset feature space. The following optimization problem is formulated [21]:

$$\min_{\mathrm{w,b,e}} J(w,e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^{N} e_i^2 \tag{2}$$

With the application of Mercer's theorem [22] for the kernel matrix Ω as $\Omega_{ij} = k(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$, $i, j = 1, \ldots, N$, it is not required to compute explicitly the nonlinear mapping $\varphi(.)$ as this is done implicitly through the use of positive definite kernel functions K [21].

$$\varsigma(w,b,e,\beta) = \frac{1}{2} w^T W + \gamma \frac{1}{2} \sum_{i=1}^{N} e_i^2 - \sum_{i=1}^{N} \beta_i \, (w^T \varphi(x_i) + b + e_i - y_i) \tag{3}$$

where β_i are Lagrange multipliers. Differentiating (3) with w, b, e_i and β_i, the conditions for optimality can be described as follow [22]:

$$\begin{cases} \dfrac{d\zeta}{dw} = 0 \to w = \sum_{i=1}^{N} \beta_i \varphi(x_i) \\ \dfrac{d\zeta}{db} = 0 \to \sum_{i=1}^{N} \beta_i = 0 \\ \dfrac{d\zeta}{db} = 0 \to \beta_i = \gamma e_i, i = 1, .. N \\ \dfrac{d\zeta}{db} = 0 \to y_i = w^T \varphi(x_i) + b + e_i, i = 1, .. N \end{cases} \tag{4}$$

By elimination of w and e_i, the following linear system is obtained [21]:

$$\left[\frac{O}{y}\right]\left[\frac{1^T}{\Omega+\gamma^{-1}I}\right]\left[\frac{b}{\beta}\right]=\left[\frac{O}{y}\right] \tag{5}$$

With $y=[y_1,\dots y_N]^T, \beta=[\beta_1,\dots\beta_N]^T$. The resulting LS-SVM model in dual space becomes:

$$y(x)=\sum_{i=1}^{N}\beta_i K(x,x_i)+b \tag{6}$$

Usually, the training of the LS-SVM model involves an optimal selection of kernel parameters and regularization parameter. For this paper, the RBF Kernel is used which is expressed as:

$$K(x,x_i)=e^{\frac{-||x-x_i||^2}{2\sigma^2}} \tag{7}$$

Note that σ^2 is a parameter associated with RBF function which has to be tuned. There is no doubt that the efficient performance of LS-SVM model involves an optimal selection of kernel parameter, σ^2 and regularization parameter, c. In [22], these parameters selection are tuned via cross-validation technique. Even though this technique seemed to be simple, the forecasting performance by using this technique is at average accuracy [23]. Thus by using ABC as an optimizer, a more accurate result is expected. In addition, ABC is known as a powerful stochastic search and optimization technique. The hybridization of ABC and LS-SVM should give better accuracy and good CTC detection.

The Artificial Bee Colony (ABC) algorithm was introduced in 2007 by Karaboga [24]. Initially, it was proposed for unconstrained optimization problems. Then, an extended version of the ABC algorithm was offered to handle constrained optimization problems [24]. The colony of artificial bees is considered as σ^2 and regularization parameter, c σ^2 and regularization parameter, c it consists of three groups of bees: employed, onlookers and scout bees. In the ABC algorithm, onlookers and employed bees with σ^2 and regularization parameter, c perform the exploitation process in the search food-source position for optimal detection of the σ^2 and regularization parameter, c results. In other hand, scouts bees with σ^2 and regularization parameter, c control the exploration process to improve CTC detection. In case of real bees, the production of new optimal σ^2 and regularization parameter, c food sources is found based on the earliest best σ^2 and regularization parameter, c results. Artificial bee with σ^2 and regularization parameter, c randomly select a foodsource position for CTC detection and produce best σ^2 and regularization parameter, c in their memory.

While, onlooker bees with σ^2 and regularization parameter, c are those bees waiting in the hive's dance area. The duration of a dance with σ^2 and regularization parameter, c is proportional to the nectar's content (fitness value) ,here error value of the classification is considered as fitness value of the optimized σ^2 and regularization parameter, c currently being exploited by the employed bee. Hence, onlooker bees watch various dances to σ^2 and regularization parameter, c and select optimal SVM parameters depending on probability proportional to the quality of that food source. The number of trials for the optimal selection of the σ^2 and regularization parameter, c is controlled by the limit value. Each cycle of the ABC algorithm comprises three steps: employed bee depending on fitness values; second, onlookers depending on probability value; third, determining the scout bees and then sends to an entirely new σ^2 and regularization parameter, c positions. In ABC algorithm creates a randomly distributed initial σ^2 and regularization parameter, c population of i solutions ($i = 1,2,\dots,E_b$), where i signifies the size of population (total number of gene samples) and E_b is the number of employed bees. Each optimal σ^2 and regularization parameter, c solution is in D-dimensional vector. The position of σ^2 and regularization parameter, c, in the ABC algorithm, represents a possible optimized σ^2 and regularization parameter, c solution .The nectar amount of a food source for σ^2 and regularization parameter, c corresponds to the error value of the classification. The Error value (Fitness value) of the randomly selected site is calculated is follows:

$$fitness_i=\frac{1}{(1+obj.Fun_i)} \tag{8}$$

Where $obj.Fun_i$ is considered as error value , After initialization, the population of σ^2 and regularization parameter, c is subjected to repeated cycles MCN , where MCN is the Maximum Cycle Number of the search process. After all employed bees, onlooker bee (O_b) evaluates the nectar σ^2 and regularization parameter, c information taken from all employed bees and chooses a food source to SVM parameters with a probability related to its nectar amount. The probability of selecting a food-source p_i by onlooker bees is calculated as follows:

$$p_i=\frac{fitness_i}{\sum_{i=1}^{E_b} fitness_i} \tag{9}$$

where fitnessi is the fitness value of a solution i, Once the new SVM parameters position is determined, another ABC algorithm cycle (MCN) starts. In other words, by adding to the current σ^2 and regularization parameter, c chosen parameter value, the neighbor food-source position is created according to the following expression:

$$x_{ij}^{new} = x_{ij}^{old} + a(x_{ij}^{old} - x_{kj}) \quad (10)$$

where k – i. The multiplier a is a random number between $[-1,1]$ and j = {1, 2, . . . ,D}. The scout produces a completely new food-source position as follows:

$$x_i^{new} = \min x_j^i + a(\max x_j^i - \min x_i^j) \quad (11)$$

where Eq. (11) applies for all j parameters. If a parameter value produced using (11) and/ or (11) exceeds its predetermined limit, the parameter can be set to an acceptable value. In this paper, the value of the parameter exceeding its limit is forced to the nearest (discrete) boundary limit value associated with it. Furthermore, the random multiplier number is set to be between [0, 1] instead of $[-1,1]$ [24].

IV.EXPERIMENTATION RESULTS

In order to perform the experimentation work we have used nine different datasets publicly available from the Gene Expression Omnibus (GEO) database [25], with their relevant characteristics. Most of the datasets provide samples from both normal and cancer breast tissues. Furthermore, there are a variety of different platforms; Affymetrix and Agilent are the most common manufacturers in this collection of datasets, while there is one dataset using a custom microarray chip from Agendia and another one from Applied Biosystems (ABI).

Each dataset that is relevant to a given comparison is downloaded from the GEO in the format (e.g., preprocessed) it has been registered. However, the raw data are not always available, and some preprocessing tasks can have already been performed. Perform k-nearest neighbor's type of imputation [26] if needed, and log-transform the probe set intensities when not already transformed. Based on the proposed hierarchical firefly algorithm methodology, have extracted the genes from the breast cancer samples, which exhibit differentially over expressed behavior. Among them all of the dataset is also used for experimentation work, but in simplification work. The first independent dataset (GSE29431) by Lopez et al. [27] provides microarray data from 65 primary breast carcinomas and 22 samples of breast normal tissues from BC patients. Considering the information about metastatic status, only include 35 tumor samples (18 metastatic, 17 nonmetastatic) and all 14 samples of breast normal tissues for validation of the proposed HFA clustering algorithm , 24 genes that identified can effectively separate the population from tumor samples. The control population shows different characteristics that enable the inclusion of most samples (nine from 12 in each test) in a single cluster. To validate the clustering results of the proposed HFA and existing hierarchical clustering algorithm for GSE29431 dataset the following metrics such as Sensitivity(Sen) ,Specificity(Spe) ,Precision(Pr) ,False Positive Rate (FPR) ,False Negative Rate (FNR) and Classification Accuracy(CA) have been used in this work.

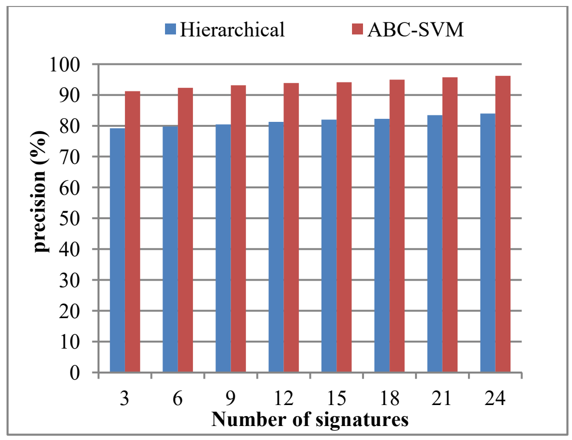


Figure 1: Precision comparison vs. ABC-SVM

The precision results of proposed hierarchical clustering and proposed ABC-SVM, so the test result shows that contribution of the work is more accurate, regardless positive is illustrated in Figure.1. Similarly precision results of proposed ABC-SVM and hierarchical clustering is defined as the percentages of predicted class which belongs to positive class, it shows that the proposed clustering methods have achieves 96.18 % and hierarchical clustering method achieves 83.98 % is illustrated in Figure.1, it is also applicable to all dataset where the resultant will be change based on the characteristics of the dataset.

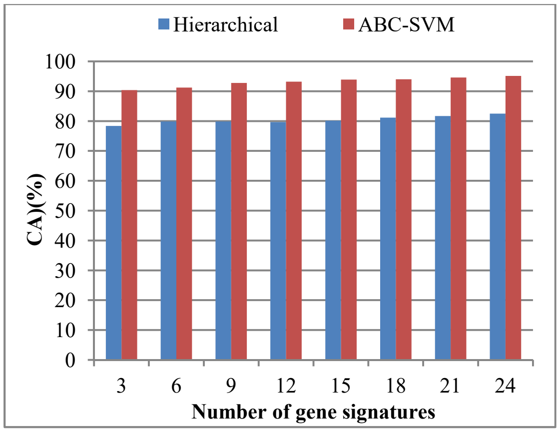


Figure 2: Accuracy comparison vs. ABC-SVM

Classification accuracy is defined as the percentage of the total amount of predictions which belongs to both positive and negative cases that were correctly identified . The accuracy results of proposed hierarchical clustering and proposed ABC-SVM, so the test result shows that contribution of the work is more accurate, regardless positive is illustrated in Figure.2. Similarly accuracy results of proposed ABC-SVM and hierarchical clustering is defined as the percentages of predicted class which belongs to positive class, it shows that the proposed clustering methods have achieves 95.12 % and hierarchical clustering method achieves 82.48 % is illustrated in Figure.2, it is also applicable to all dataset where the resultant will be change based on the characteristics of the dataset.

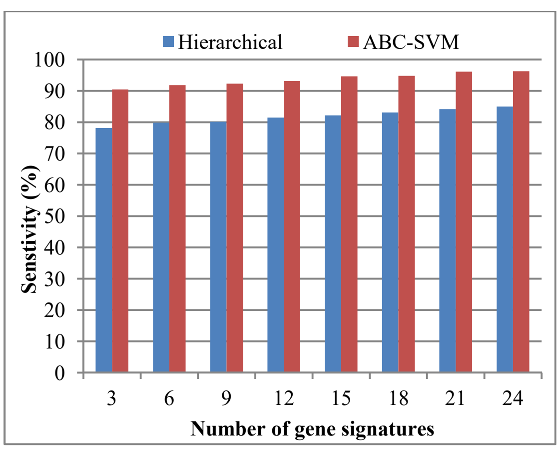


Figure 3: Sensitivity comparison vs. ABC-SVM

The sensitivity results of proposed ABC-SVM and Hierarchical clustering represents the percentage of actual true positive results for GSE29431 dataset samples to identify the CTC and detect the CTC in BC. Sensitivity results of posed ABC-SVM clustering is 96.28 % and Hierarchical clustering achieves 84.98 % clustering, so the test result shows that contribution of the work is more accurate, regardless positive is illustrated in Figure.3. It is also applicable to all dataset where the resultant will be change based on the characteristics of the dataset.

V.CONCLUSION

Circulating Tumor Cells (CTC) in the blood tissue plays a critical role in establishing metastases. In this paper, we describe a hybrid Artificial Bee Colony (ABC) approach that attempts to explore the field by combining microarray gene expression data originated from tissue and PB. The ABC algorithm is used to obtain the optimal values of regularization parameter c and Kernel RBF parameter, σ^2, which are embedded in LS-SVM toolbox and adopt a supervised learning approach to LS-SVM model for the identification and characterization of CTC. It also shows that the proposed ABC-SVM results are compared to existing method and it has been proved and achieved best results. The proposed ABC-SVM result achieves results in terms of their association with CTC assessing their potential for direct identification of CTC cells and express EMT markers.

REFERENCES

[1] Howard, E.W., S.C. Leung, H.F. Yuen, C.W. Ch ua, D.T. Lee, K.W. Chan, X. Wang,Y.C. Wong. 2008. Decreased adhesiveness, resistance to anoikis and suppression of GRP94 are integral to the survival of circulating tumor cells in prostate cancer. Clin. Exp. Metastasis.25:497–508.

[2] Hüsemann, Y., J.B. Geigl, F. Schubert, P. Musiani, M. Meyer, E. Burghart, G. Forni, R. Eils, T. Fehm, G. Riethmüller, C.A. Klein. 2008. Systemic spread is an early step in breast cancer. Cancer Cell. 13:58–68

[3] Tsuji, T., S. Ibaragi, K. Shima, M.G. Hu, M. Kat surano, A. Sasaki, G.F. Hu. 2008. Epithelial-mesenchymal transition induced by growth suppressor p12CDK2-AP1 promotes tumor cell local invasion but suppresses distant colony growth. Cancer Res. 68:10377–10386

[4] Tsuji, T., S. Ibaragi, G.F. Hu. 2009. Epithelial-mesenchymal transition and cell cooperativity in metastasis. Cancer Res. 69:7135–7139

[5] Kaplan, R.N., R.D. Riba, S. Zacharoulis, A.H. Bramley, L. Vincent, C. Costa, D.D. MacDonald, D.K. Jin, K. Shido, S.A. Kerns, et al. 2005. VEGFR1-positive haematopoietic bone marrow progenitors initiate the pre-metastatic niche. Nature. 438:820–827.

[6] Kim, M.Y., T. Oskarsson, S. Acharyya, D.X. Nguyen, X.H. Zhang, L. Norton, J. Massagué. 2009. Tumor self-seeding by circulating cancer cells. Cell. 139:1315–1326.
[7] J. Barbaz´an, L. Alonso-Alconada, L. Muinelo-Romay, M. Vieito, A. Abalo, M. Alonso-Nocelo, S. Candamio, E. Gallardo, B. Fern´andez, I. Abdulkader, M. de Los A´ ngeles Casares, A. Go´mez-Tato, R. Lo´pez- L´opez, and M. Abal, "Molecular characterization of circulating tumor cells in human metastatic colorectal cancer," PloS One, vol. 7, no. 7, p. e40476, 2012.
[8] E. Obermayr, F. S. Cabo, M. K. Tea, C. Singer, M. Krainer, M. Fischer, J. Sehouli, A. Reinthaller,R.Horvat, G. Heinze, D. Tong, andR. Zeillinger, "Assessment of a six gene panel for the molecular detection of circulating tumor cells in the blood of female cancer patients," BMC Cancer, vol. 10, no. 1, p. 666, 2010.
[9] T. J. Molloy, P. Roepman, B. Naume, and L. J. V. Veer, "A prognostic gene expression profile that predicts circulating tumor cell presence in breast cancer patients," PloS One, vol. 7, no. 2, p. e32426, Feb. 2012.
[10] Marrinucci, D., K. Bethel, R.H. Bruce, D.N. Curry, B. Hsieh, M. Humphrey, R.T. Krivacic, J. Kroener, L. Kroener, A. Ladanyi, et al. 2007. Case study of the morphologic variation of circulating tumor cells. Hum. Pathol. 38:514–519.
[11] Pachmann, K., J.H. Clement, C.P. Schneider, B. Willen, O. Camara, U. Pachmann, K. Höffken. 2005b. Standardized quantification of circulating peripheral tumor cells from lung and breast cancer. Clin. Chem. Lab. Med. 43:617–627
[12] Cristofanilli, M., J. Reuben, J. Uhr. 2008. Circulating tumor cells in breast cancer: fiction or reality? J. Clin. Oncol. 26:3656–3657.
[13] Botteri, E., M.T. Sandri, V. Bagnardi, E. Munzone, L. Zorzino, N. Rotmensz, C. Casadio, M.C. Cassatella, A. Esposito, G. Curigliano, et al. 2010. Modeling the relationship between circulating tumour cells number and prognosis of metastatic breast cancer. Breast Cancer Res. Treat. 122:211–217.
[14] Wülfing, P., J. Borchard, H. Buerger, S. Heidl, K.S. Zänker, L. Kiesel, B. Brandt. 2006. HER2-positive circulating tumor cells indicate poor clinical outcome in stage I to III breast cancer patients. Clin. Cancer Res. 12:1715–1720.
[15] Pierga, J.Y., F.C. Bidard, C. Mathiot, E. Brain, S. Delaloge, S. Giachetti, P. de Cremoux, R. Salmon, A. Vincent-Salomon, M. Marty. 2008. Circulating tumor cell detection predicts early metastatic relapse after neoadjuvant chemotherapy in large operable and locally advanced breast cancer in a phase II randomized trial. Clin. Cancer Res. 14:7004–7010.
[16] Lin HK, Zheng S, Williams AJ, Balic M, Groshen S, Scher HI, et al. Portable filter-based microdevice for detection and characterization of circulating tumor cells. Clin Cancer Res 2010;16:5011– 8.
[17] Zheng S, Lin HK, Lu B, Williams A, Datar R, Cote RJ, Tai YC. 3D microfilter device for viable circulating tumor cell (CTC) enrichment from blood. Biomed Microdevices 2011;13:203–13.
[18] V. G. Tusher, R. Tibshirani, and G. Chu, "Significance analysis of microarrays applied to the ionizing radiation response," Proc. Nat. Acad. Sci., vol. 98, no. 9, pp. 5116–5121, Apr. 2001.
[19] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," J. Roy. Statist. Soc. Series B, vol. 57, pp. 289–300, 1995
[20] Espinoza M, Suykens J, Moor B. Fixed-size least squares support vector machines: a large scale application in electrical load forecasting. Comput Manage Sci 2006;3:113–29.
[21] Suykens JAK, Gestel TV, De Brabanter J, De Moor B, Vandewelle J. Least squares support vector machines. Singapore: World Scientific; 2002.
[22] Espinoza M, Suykens JAK, Belmans R, De Moor B. Electric load forecasting. Control Syst Mag, IEEE 2007;27:43–57.
[23] Lean Y, Huanhuan C, Shouyang W, Kin Keung L. Evolving least squares support vector machines for stock market trend mining. Evol Comput, IEEE Trans 2009;13:87–102.
[24] Abu-Moutiand FS, El-Hawary ME. Modified artificial bee colony algorithm for optimal distributed generation sizing and allocation in distribution systems. IEEE electrical power and energy conference (EPEC); 2009.
[25] R. Edgar, M. Domrachev, and A. E. Lash, "Gene expression omnibus: NCBI gene expression and hybridization array data repository," Nucl. Acids Res., vol. 30, no. 1, pp. 207–210, Jan. 2002.
[26] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, and R. B. Altman, "Missing value estimation methods for DNA microarrays," Bioinformatics, vol. 17, no. 6, pp. 520–525, Jun. 2001.
[27] F. J. Lopez, M. Cuadros, C. Cano, A. Concha, and A. Blanco, "Biomedical application of fuzzy association rules for identifying breast cancer biomarkers," Med. Biol. Eng. Comput., vol. 50, no.9 pp. 984 – 990, Sep. 2012.

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 3, March 2016

Moving Object Segmentation and Vibrant Background Elimination Using LS-SVM

[1]Mehul C. Parikh, [2]Kishor G. Maradia

[1]Research Scholar, Computer Engineering Department, Chartoar University of Science and Technology Changa,Gujarat,India.

[2]Professor, Department ofElectronics and Communication,Government Engineering College, Gandhinagar,Gujarat,India.

Abstract: Moving object segmentation is a significant research area in the field of computer intelligence due to technological and theoretical progress. Many approaches are being developed for moving object segmentation. These approaches are useful for specific situation but have many restrictions. Execution speed of these approaches is one of the major limitations. Machine learning techniques are used to decrease time and improve quality of result. LS-SVM optimizes result quality and time complexity in classification problem. This paper describes an approach to segment moving object and vibrant background elimination using the least squares support vector machine method. In this method consecutive frame difference was given as a input to bank of gabor filter to detect texture feature using pixel intensity. Mean value of intensity on 4 * 4 block of image and on whole image was calculated and which are then used to train LS-SVM model using random sampling. Trained LS-SVM model was then used to segment moving object from the image other than the training images. Results obtained by this approach are very promising with improvement in execution time.

Key Words: Segmentation, Machine Learning, Gabor filter, LS-SVM.

I. INTRODUCTION

Segmentation process is to classify the semantically meaningful elements of an image and grouping the pixels belonging to such components. Motion segmentation is the grouping of pixels that are associated with a smooth and uniform motion profile. In the recent years, there are an extensive range of very interesting and innovative methods in the collected works of the moving object segmentation, and these methods can be approximately classified into the following categories: image difference thresholding based, optical flow based, statistical based on motion estimation, 2D approach, 3D approach, wavelet based, clustering based, genetic algorithm based and machine learning based [12]. This partition is not mean tight and few algorithms can be cited in more than single class. Important attribute of motion segmentation algorithm are feature based, dense based, occlusion, multiple object, spatial continuity, robustness, computation time and complexity. Segmentation processes with different approaches are implemented in specific condition with predefined parameter which gives cost and time effective result.

However, these approaches are not giving satisfactory result in condition like change in camera position, lighting condition, object location, etc.

Motion base object segmentation is a non-polynomial hard problem. Machine learning algorithms are current research paradigm. These algorithms are used to solve many non-polynomial hard problems with less complexity. Support vector machine is one of the approach which is mostly used for classification. Motion base object segmentation can also be observed as classification problem. Moving and non moving pixels classified into two dissimilar classes and SVM gives prominent results in these approaches. Least squares support vector machine (LS-SVM) is a novel kind of SVM, which are a set of related supervised learning methods that analyzes data and distinguish patterns, and which are used for classification and regression analysis. This approach considerably decreases the complexity and the computation time. In this research article LS-SVM method used for, removing dynamic background and segment moving object. The section two presents theoretical backgrounds, section three describes basic theory of LS-SVM, section four represents propose algorithm for moving object segmentation and dynamic background removal, section five presents simulation results and section six concludes the paper.

II. Theoretical Backgrounds

Image difference and thresholding [1]-[5]: It is the simplest and most commonly used technique for detecting change. Two consecutive frames are compared pixel by pixel for some fixed threshold value; the result of which indicates temporal changes. Frame difference and background subtraction are basic simplest image difference technique. Many researchers have used the combination of these two methods for background modeling and background updating using median filter. Frame difference output have also been used with edge detection operator canny and sobel to get edge of moving object. Some researcher have also used histogram to generate background from series of frame and a foreground was detected by comparing each frame with given background of predefined threshold.

Statistical based [6]-[8]: Use of statistical method is widely found in the literature of unraveling motion segmentation problem. Instead of thresholding, this method compares statistical performance of trivial areas at each pixel position in the consecutive frames. Kalman filter has also been used for a prediction and correction of pixel value to detect foreground and background. The gray histogram entropy has also been employed with statistical property of the motion edge with canny operator.

Optical flow based [9]-[11]: Optical flow is transitory speed field which is prepared by moving pixels of moving object surface in space. Optical flow reflects the image alterations due to motion for the period of a time interval ∂_t. As optical flow is abstraction, it represents only those motion related intensity changes in the image that required in further processing. One of the researcher used variation in original Lucas Kanade algorithm to detect moving object. Motion vector of vibrant background and moving object region were figured by one of the researcher and from this information motion histogram was prepared and updated adaptively according to motion information which was used to detect the moving object. Horn and Schunck algorithm has also been used by a researcher to calculate optical flow and then uses a gamma distribution was used to label moving and stationary pixel to detect moving object.

Clustering based [13]: In this type of algorithm each pixel is classified by assemblage of K clusters where each cluster consists of a weight and an average pixel value of centroid C_k . Incoming frame pixel are compared with the corresponding cluster group. The matching cluster with the highest weight is searched such that manhattans distance between its centroid and the incoming pixel is below to user prescribed threshold T .

Region based [14]-[15]: Image is classified into a number of regions or classes in this method. Each pixel in the image, need to decide that it belongs to which class or region, subsequently attention base region growing algorithm extracts object displacement between frames by comparing salient region and then region growing classifies motion region according to motion information.

Genetic algorithm based [16]: Genetic algorithm uses the key relevance of video images to expedite the evolutionary development and match with uncertain evolutionary process. The accurate video segmentation has been achieved with low computational complexity using this algorithm.

These algorithms were tried to achieve success in several applications, however none of them are typically applicable to any or all form of moving object scenario. Several approaches and their corresponding enhancements have been planned to confirm the accuracy and time efficiency of motion segmentation. However, lot of works needs to done to beat their drawbacks, and new method needs to be developed using alternative domains, particularly machine learning. Motion segmentation may be viewed as a classification problem constructed on texture features. Recently, intellectual approaches, like neural network and support vector machine (SVM) have already been used with success in image segmentation.

LaetitiaLeyrit[17] proposed adaptive boosting algorithm for features selection and AdaBoost to selected a subset of them as binary vector in a kernel based machine learning classifier. Kwang-Bake kin [18] proposed method that uses sobel operator for edge detection and noise removal. This has been used for background removal, ART2 based hybrid network architecture with RBF kernel used in middle layer neuron and sigmoid function in output layer neurons. Shih-Chia Huang[19] suggested pyramidal background matching structure for motion detection. Noise was removed using Bezier curve then probability mass function and cumulative distribution function were used to calculate threshold value to generate binary motion mask.

QingsongZhu[20] proposed novel recursive bayesian learning method, which uses multilayer gaussian distribution function for construction of background. This background was updated via recursive bayesian estimation. Foreground was obtained by deducting this updated background frame by frame. Cui Liang [21] proposed a moving vehicle segmentation method using semi fuzzy cluster algorithm with edge base information. Every edge pixels will be associated into the most reasonable region according to the semi fuzzy cluster algorithm. Finally, the region that is similar with the background will be detected and the remaining regions are considered as moving vehicle within the frame.

KeyvanKasiri[22] proposed hierarchical method for brain segmentation using atlas information and LS-SVM was used to generate brain tissue probabilities. Quantitative and qualitative results of their simulations demonstrate excellent performance of the applied method in segmenting brain tissues. HaiyanZhao[23] suggested LS-SVM based character classification method for license plate recognition system. Their result shows that recognition time is reduced drastically and in 19.4 ms one character is recognize. JianhongXie [24] proposed LS-SVM method to classify optical character in optical character recognition method. Their result also shows that accuracy increases and time complexity also gets reduced.

Hong Ying Yang [25] proposed LS-SVM based image segmentation method using color and texture information. They selected the HSV color space to extract pixel level color feature and gabor filter to extract the texture feature of the image. The arimoto entropy method was used to select the samples.LS-SVM model was trained using this two features and image is segmented through LS-SVM classification. Proposed algorithm achieves better quantitative results. Hence, LS-SVM appears as a powerful supervised learning method with high generalization characteristics.

For quality result of motion base object segmentation, numerous researchers evaluated various properties of video clip by difficult formulae which increase time complexity and scope of algorithm is specific to application or scenario. A multipurpose algorithm for motion base object segmentation requires to be developed. In this paper, competent motion based moving object segmentation algorithm using texture aspect with LS-SVM model is presented.

III. THEORETICAL SUPPORT [26]-[28]

Machine learning process can be described as development of algorithm that naturally enhance with experience and implementing a learning process. Machine learning algorithm can be classified into following types: supervised learning, unsupervised learning, semi supervised learning, reinforcement learning. Linear function is the simplest form of separation. Linear function $f(x)$ for separation can be written as $f(x)=\left[w^T+b\right]$, where, w is the weight vector and b as bias. Vapnik and Chervonenkis hypothesize that the generalization ability depends on distance between hyper plane and the training points. They presented the generalize depiction, a learning algorithm for separable problems. They constructed a hyper plane which maximally separates the classes. The separating hyper plane described as w and b. For construction of the optimal hyper plane, support vector machine formulates the problem in primal weight as constrained optimization problem as shown below:

$$\min_{w,b} J_p\left(w\right)=\frac{1}{2}w^T w \ such\,that\, y_k\left[w^T\ x_k+b\geq 1,k=1,...N\right] \tag{1}$$

The Lagrangian for this Problem is

$$L\left(w,b;\alpha=\frac{1}{2}w^T w-\sum_{k=1}^{N}\alpha_k\left(y_k\left[w^T x_k+b\right]-1\right)\right. \tag{2}$$

Resulting classifier is:

$$y(x)=sign\left[\sum_{k=1}^{N}\alpha_k y_k x_k^T x+b\right] \tag{3}$$

SVM Classifier in dual space takes the from

$$\max_{\alpha} J_D(\alpha)=-\frac{1}{2}\sum_{k,l=1}^{N} y_k y_l x_k^T\, x_l\, \alpha_k\alpha_l+\sum_{k=1}^{N}\alpha_k\ suchthat\sum_{k=1}^{N}\alpha_k y_k=0, \alpha_k\geq 0 \tag{4}$$

SVM calculates the optimal separating hyper plane in the feature space. Optimal separating hyper plane defined as the maximum margin hyper plane in the higher dimensional feature space.

Least squares support vector machine (LS-SVM) is new kind of SVM, which was proposed by Suykens and Vandewalle. LS-SVM is the part of kernel-based learning method category. The computation speed of this algorithm is faster than the other SVM. In this form, the solution can be found by solving a set of linear equations instead of a convex quadratic programming (QP) problem for classical SVM. Change suggested by Suyken is shown below:

$$\min_{w,b,e} J_p\left(w,e\right)=\frac{1}{2}w^T w+\gamma\frac{1}{2}\sum_{k=1}^{N}e_k^2\ suchthat\ y_k\left[w^T\,\varphi(x_k)+b=1-e_k, k=1,...N\right] \tag{5}$$

Classifier in the primal space takes the form

$$y(x)=sign\left[w^T\,\varphi(x)+b\right] \tag{6}$$

The vapnik formulation is modified at two points 1) equality constraint is used. 2) For Error variable e_k a square loss function is taken. The Lagrangian for the problem is:

$$L(w,b,e;\alpha)=J_p\left(w,e\right)-\sum_{k=1}^{N}\alpha_k y_k\left[w^T\,(\varphi x_k)+b\right]-1=e_k \tag{7}$$

The classifier in the dual space takes the form

$$y(x)=sign\left[\sum_{k=1}^{N}\alpha_k y_k\, K\left(x,x_k\right)+b\right] \tag{8}$$

Where, α_k values are Lagrange multiplier, which can be positive or negative and $K\left(x;x_k\right)$ is the kernel trick. LS-SVM simplifies the SVM formulation by replacing inequality constraint to equality constraint. This helps in reducing the complexity and the computation time significantly.

IV. Gabor Filters For Texture Feature Extraction [28-32]

Two dimensional gabor filter was proposed by Dauman to model the spatial summation properties of simple cell in the visual codex. They are widely used in image processing for extraction of texture features. Basic function of two dimensional gabor filter is:

$$Q_{\xi\eta\lambda\Theta\varphi}(X,Y)=\exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{X'}{\lambda}+\varphi\right) \tag{9}$$

Where, x and y argument specify the location of a pixel. Wavelength (λ) is the cosine factor of the gabor filter kernel. Value of wavelength is specified in pixels. Its value is only real number and equal or greater than 2. To avoid undesired effects at the image borders, its value should be smaller than one fifth of the input image size. Orientation (θ) is normal to the parallel stripes of gabor function. Its value is specified in degree. Phase offset (φ) used as argument of the cosine factor of the gabor function. 0^o and 90^o are considered in this approach. Spatial aspect ratio is defined by γ . Bandwidth (B) of a gabor filter related to the ratio $\frac{\sigma}{\lambda}$. The value of σ cannot be specified directly. It can be changed through the bandwidth b. Frame difference of input frame sequence given to the bank of eight gabor filter. The output of this gabor filter shows that dynamic background pixel value is smaller than the moving object pixel value. Hence, there is a scope of classifying the pixel as background and foreground using LS-SVM classifier.

V. Ls-Svm Moving Object Segmentation Using Texture Information [28]-[32]

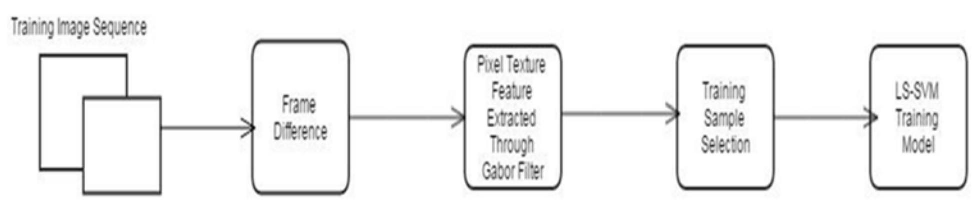


Figure. 1: Block diagram of LS-SVM Training Model.

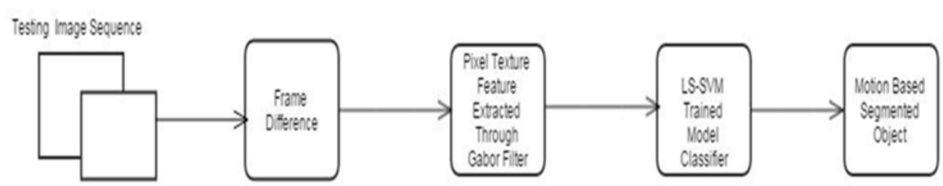


Figure. 2: Block diagram of LS-SVM Testing Model.

First, frame difference of input image sequence was calculated then this difference values were given to gabor filter bank. Here, eight gabor filter were used and selection of various parameter are as under:

Wavelength (λ) : 3 and 8

Orientation (θ) : 0

Phase Offset (φ) : 0, 90

Aspect ratio (γ) : 0.5 and 0.75

Bandwidth (b) : 1

Texture feature of this image was selected by converting this image to 4*4 blocks, and mean value of each block were calculated. Mean value of whole image was also calculated. Similarly, 5 image sequences are used to extract texture feature. From this randomly 4800 samples are selected for training purpose. Both features were then given as input to LS-SVM training model. If the mean value of image block was obtained greater than T_{tr} then block was considered as positive support vector and assigned $as+1$ otherwise -1. T_{tr} is a threshold value. The RBF kernel was used for this algorithm. Training time was approximately 30 minute which varies with the type of image and environment variables. Value of γ = 3.7957 and σ_2 = 0. 039642 obtained by training. For testing purpose first three steps used for training LS-SVM model were adopted. Output of the pixel level feature block given to LS-SVM classifier and finally moving objects were segmented for the given image sequence.

VI. Experiments And Results

The proposed algorithm was implemented using Matlab R12. It was run on a Sony personal computer, using a 2.3 GHz core i3 processor with a 4GB Random Access Memory. For this simulation, the KULeuven's LS-SVMlab MATLAB/C [33] toolbox was employed to handle the training and testing techniques. In the learning algorithm, radial basis function (RBF) was chosen as the kernel function of LS-SVM. Testing image of a single car sequence, shown in figure 3(a), (b) [34]. Other sequences are also given for training and these sequences were passing from gabor filter bank. Output of gabor filter demonstrated in figure 3(c). Algorithm was tested on nine sequences.

(a) (b)

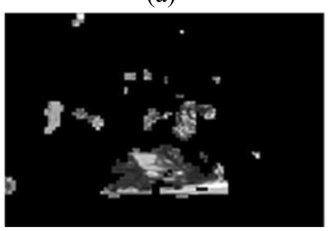
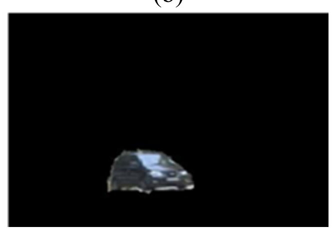
(c)

Figure. 3: (a) (b) Original test images (c) gabor filter output

The test image files were from caviar project and I2R dataset [35] - [36]. Testing time of these sequences found around 3 second. Threshold value 0.30 is empirically selected based on experiments.

(a) (b)

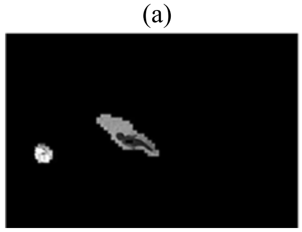
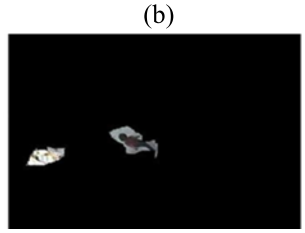
(c) (d)

Figure. 4 – Campus Sequence: (a) and (b) Original Sequence (c) Segmented Output (d) Ground truth

Figure 4 (a) and (b) shows campus sequence from I2R dataset [36]. This sequence contains total two thousand four hundred thirty eight images. Algorithm tested on hundred and sixteen images from total images of this sequence. One of the results displayed in above figure. Background is very dense with all weaving trees. Black color moving car is detected. Few leaves of trees are also detected.

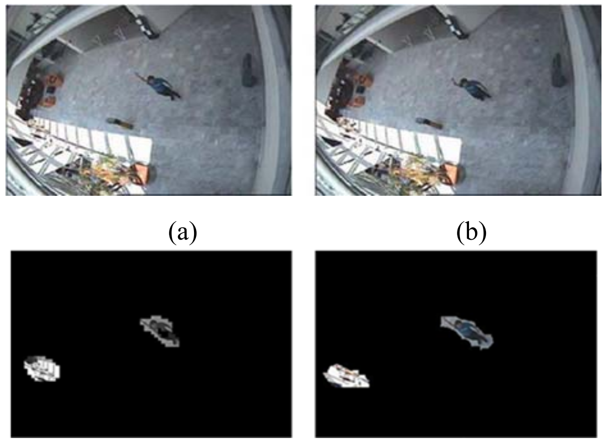
(a) (b)

(c) (d)

Figure. 5 Walk Sequence -1: (a) and (b) Original Sequence (c) Segmented Output (d) Ground truth

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 3, March 2016



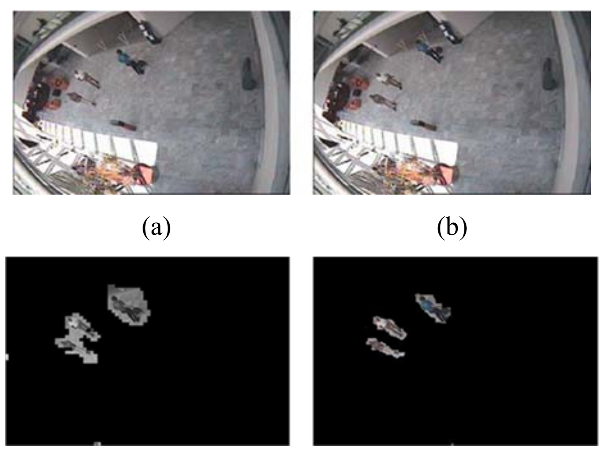
(a) (b)

(c) (d)

Figure. 6 Walk Sequence -2: (a) and (b) Original Sequence (c) Segmented Output (d) Ground truth

Figure 5 and 6 shows walk sequence from CAVIAR project dataset [35]. This sequence contains total one thousand six hundred ten images. Algorithm tested on about two hundred images from total images of this sequence. Figure 5 (a) and (b) shows lady is moving and some movement is in sunny environment near window. Figure 6 (a) and (b) shows a person is raising his hand and some movement is in sunny environment near window. All moving objects detected perfectly in both sequence. Some part of sunny environment is detected in both sequences.

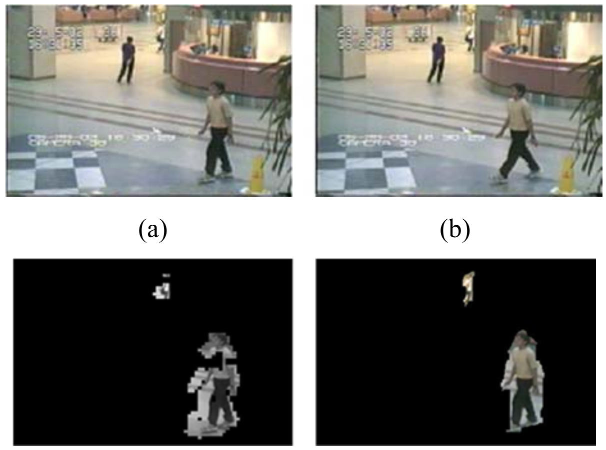
Figure 7 shows left bag sequence from CAVIAR project dataset [35]. This sequence contains total one thousand four hundred thirty eight images. Algorithm tested on about one hundred images from total images of this sequence. Figure 7 (a) and (b) shows three persons are moving. Figure 7(c) shows that all moving persons detected perfectly.



(a) (b)

(c) (d)

Figure.7 Left Bag: (a) and (b) Original Sequence (c) Segmented Output (d) Ground truth



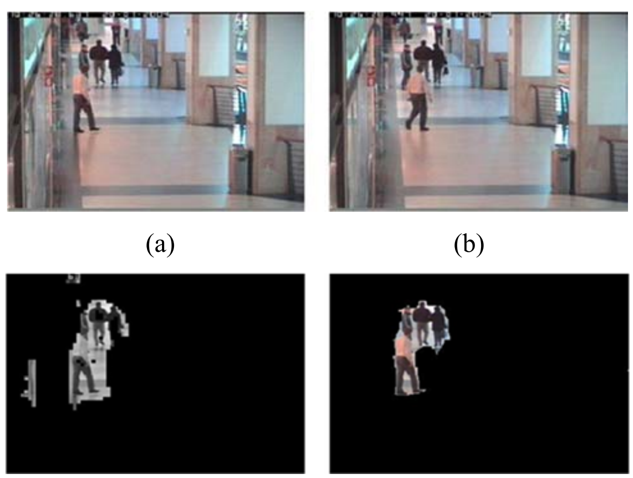
(a) (b)

(c) (d)

Figure.8 Air-Port Sequence : (a) and (b) Original Sequence (c) Segmented Output (d) Ground truth

Figure 8 shows Airport sequence from I2R dataset [36]. This sequence contains total four thousand five hundred and eighty three images. Algorithm tested on about one hundred and thirty images from total images of this sequence. Figure 8 (a) and (b) shows that one person is moving in front of a tree. One person is standing in middle and in upper part one another person is moving. All moving persons are detected. All stationary objects are removed perfectly.

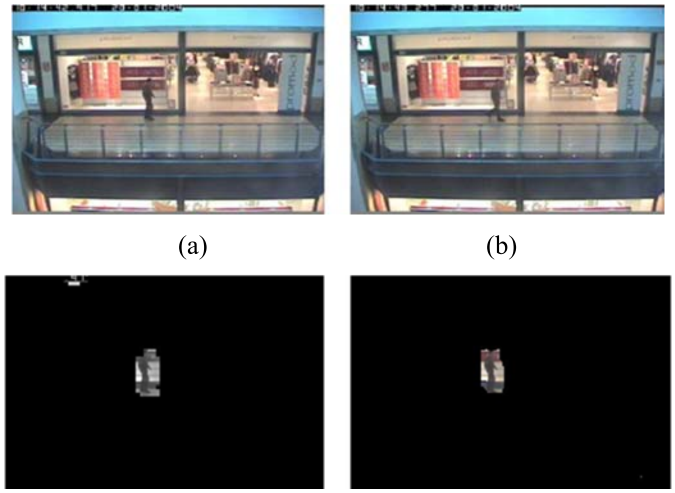
Figure 9 shows one leave shopping corridor sequence from CAVIAR project dataset [35]. This sequence contains total two hundred and ninety four images. Algorithm tested on about twenty images from total images of this sequence. Figure 9 (a) and (b) shows one person is leaving from shop and moving in corridor and three persons are moving in corridor. All moving persons are detected perfectly.



(a) (b)

(c) (d)

Figure.9 Lobby Sequence: (a) and (b) Original Sequence (c) Segmented Output (d) Ground truth

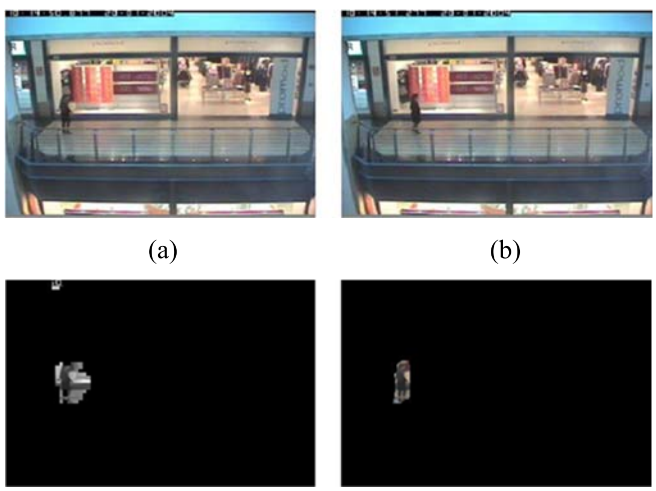


(a) (b)

(c) (d)

Figure.10 Shopping Mall Sequence: (a) and (b) Original Sequence (c) Segmented Output (d) Ground truth

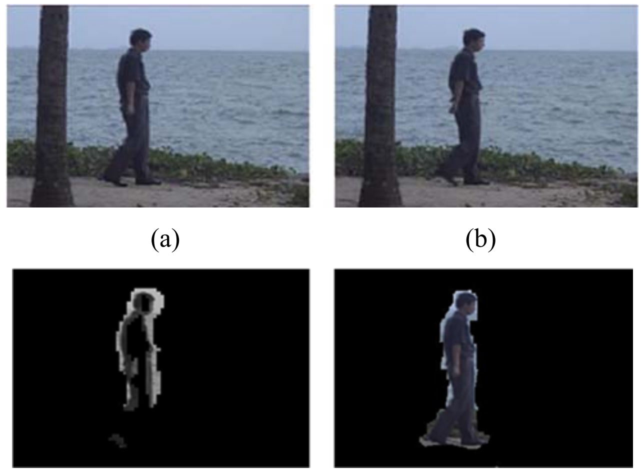
Figure 10 and 11 shows one stop no enter sequence from CAVIAR project dataset [35]. This sequence contains total seven hundred and twenty four images. Algorithm tested on about sixty images from total images of this sequence. Figure 10- 11(a) and (b) shows a person and a lady are moving respectively. From both of sequence moving person and lady detected completely.



(a) (b)

(c) (d)

Figure.11 Shopping Mall Sequence - 2: (a) and (b) Original Sequence (c) Segmented Output (d) Ground truth



(a) (b)

(c) (d)

Figure.12 Water Surface Sequence: (a) and (b) Original Sequence (c) Segmented Output (d) Ground truth

Figure 12 shows water surface sequence from I2R dataset [36]. This sequence contains total one thousand six hundred and thirty two images. Algorithm tested on about sixty six images from total images of this sequence. Figure 12 (a) and (b) shows one persons is moving in the front of sea. Here, water of sea is also moving. Segmented output shows the person is detected but some part of person as background. All the moving sea water is removed totally.

VII. Quantitative Evaluations And Computational Cost

First hand base segmented ground truth is prepared. Each segmented output was compared with ground truth. All sequences were evaluated using false positive ratio and true positive ratio as per below mentioned table -1.

TABLE – 1

QUANTITATIVE EVALUATION

Sequence	***False Positive Ratio***	***True Positive Ratio***
Campus	0.0678	0.7990
Walk Sequence - 1	0.0100	0.7287
Walk Sequence - 2	0.0089	0.5638
Left Beg	0.0242	0.9518
Airport	0.0249	0.7747
Lobby	0.0292	0.8113
Shopping Mall - 1	0.0056	0.8920
Shopping Mall - 2	0.0106	0.8360
Water Surface	0.0143	0.4004

T.P.R values in most of sequence are above 0.75 except walk sequence -2 and water surface sequence. It shows that the segmented output is matching with ground truth. Testing time is near to 3 seconds. This result indicates that this algorithm can work in real time as testing time is very less.

VII. CONCLUSION

A new approach for moving object segmentation and vibrant background removal using least squares support vector machine is introduced in this presentation. The algorithm used is based on pixel classification with its local information intensity and the generalized ability of LS-SVM classifier is utilized. It is observed that results of walk sequence, left bag sequence, airport, one stop no enter gives mostly perfect foreground detection and vibrant background removal. It is also seen that in every test sequence utmost of vibrant background is removed. Thus, results demonstrate that it is working for indoor and outdoor type sequences. Selection of threshold is a measure concern to improve the performance of this algorithm which right now decided based on experiments. Future work may be focused on development of adaptive threshold selection expecting improvement in the performance of stated algorithm.

ACKNOWLEDGEMENT

We are greatly thankful to Dr. Amit Ganatara, Dean, Charotar University of Science and Technology, for giving opportunity for research.

REFERENCES

[1] Shao-Yi Chien.,Shyh-Yih Ma, and Liang-Gee Chen, ”Efficient Moving Object Segmentation Algorithm using Background Registration Technique”,IEEEtranscation on circuit and system of video technology, 12 (7), July-2002,pp.577-586.

[2] Shrieen Y. Elhabian, Khaled M. El-Sayed and Sumaya H. Ahmed, ”Moving object detection in spatial domain using background removal techniques Sate of Art”. Recent Patents on Computer Science, Bentham Science Publishers Ltd.,1,2008,pp.32-54.

[3] Jie Cao, Li Li, ”Vehicle Objects Detection of Video Image Based on Gray Scale Characteristics”, in First International workshop on Education technology and computer science, March-2009, pp.936-940.

[4] Jie He; KebinJia; ZhuoyiLv, ”A Fast and Robust Algorithm of Detection and Segmentation for Moving Object,” Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2009. IIH-MSP ’09.,pp.718,721, 12-14 Sept. 2009 doi: 10.1109/IIH-MSP.2009.74

[5] Nicholas A. Mandellos, Iphigenia Keramitsoglou, Chris T. Kiranoudis, ”A background subtraction algorithm for detecting and tracking vehicles”, Elsevier-Expert system with applications,Volume 38, Issue 3, March 2011, pp. 1619-1631

[6] Liu Liu; Jian-Xun Li, ”A Novel Shot Segmentation Algorithm Based on Motion Edge Feature,” Symposium onPhotonics and Optoelectronic (SOPO), 19-21 June 2010,pp.1,5,doi: 10.1109/SOPO.2010.5504194

[7] Ling Shao, Ling Ni, Yan Liu, Jianguo Zhang, ”Human action segmentation and recognition via motion and shape analysis”, Elsevier-Pattern Recognition Letters,33,2012,pp.438-445.

[8] SaeedKamel,HosseinEbrahimnezhad, AfshinEbrahimi.,”Moving Object Removal in Video Sequence and Background Restoration using Kalman Filter”, in International Symposium on Telecommunications, 27-28 Aug-2008, pp.580-585.

[9] Wei Zhang, Q.M. Jonathan Wu, Haibing Yin, “Moving vehicles detection based on adaptive motion histogram”,Elsevier-Digital Signal Processing 20 ,2010,pp793-805.

[10] Cheolkon Jung, L.C. Jiao, Maoguo Gong, ”A novel approach to motion segmentation based on gamma distribution”, Elsevier- International Journal of Electronics and Communication,Vol.66,Issue 3,March 2012,pp.235-238.

[11] Lee Yee Siong; Mokri, S.S.; Hussain, A.; Ibrahim, N.; Mustafa, M.M., ”Motion detection using Lucas Kanade algorithm and application enhancement,” International Conference on Electrical Engineering and Informatics,Selangor, Malaysia, 5-7, Aug. 2009. ICEEI ’09., vol.02, pp.537,542, doi: 10.1109/ICEEI.2009.5254757

[12] Luca Zappella, Xavier Llado, JoaquimSalvi,”New Trends in Motion Segmentation”, Pattern Recognition book. In-TECH, Ed. Intechweb.org. ISBN: 978-953-307-014-8,2009,pp.31-46.

[13] Jong Bae Kim, Hang JoonKim.,”Efficient region based motion segmentation for a video monitoring system”, Elsevier - Pattern recognition letters 24,2003,pp.113-128.

[14] ShijieZhang,FredStentiform, ”Motion Segmentation using Region Growing And An Attention Based Algorithm”, 4th European Conference on Visual Media Production, 27-28 Nov.2007. IETCVMP, pp.1-8.

[15] Ying-Tung Hsiao, Cheng-Long Chuang, Yen-Ling Lu, Joe-Air Jiang., ”Robust multiple objects tracking using image segmentation and trajectory estimation scheme in video frames”, Elsevier - Image and Vision Computing 24,2006, pp 1123-1136.

[16] Yan-ni WANG, Yang-yu Fan, ”Adaptive Motion Segmentation Based on Genetic Algorithm”, IEEE, 2nd International Congress on Image and Signal Processing,17-19, Oct, 2009, Tianjin. CISP ’09.pp 1-4.

[17] LaetitiaLeyrit, Chateau and Jead Thierry Lapreste,”Classifiers Association for High Dimensional Problem : Application to Pedestrian Recognition”, New Advances in Machine Learning - Intech Publication, 2010, pp.93-102.

[18] Kwang-Baeak Kim, Sungshim Kim and Young Woon Woo, ”An Intelligent System for container Image Recognition using ART2 based Self Organizing Supervised Learning Algorithm”, New Advances in Machine Learning - Intech Publication, 2010,pp.163-171.

[19] Shih-Chia Huang, Fan-Chieh Cheng, ”Motion detection with pyramid structure of background model for intelligent surveillance systems”, Elsevier - Engineering Applications and Artificial Intelligence, Volume 25,2012, pp.1338-1348.

[20] Qingsong Zhu, Zhan Song, YaoqinXie and Lei Wang, ”A Novel Recursive Bayesian Learning Based Method for the Efficient and Accurate

Segmentation of Video with Dynamic Background", IEEE Transaction on Image Processing, 21(9), 2012 ,pp.3865-3876.

[21] Cui Liang, Wang Xingepeng, FengGuojie, "Segmentation Method of Moving Vehicles Based on Semi-fuzzy Cluster", 3rd International Conference on Software Engineering and Service Science (ICSESS), 2012, pp. 147-149,

[22] KeyvanKasiri, Kamran Kazemi, MohammandJavadDehghani, Mohammad SadeghHelfroush, "Atlas based segmentation of Brain MR image using Least Squares Support Vector Machines", 2nd International Conference on Image Processing Theory Tools and Applications (IPTA), 7-10 July 2010 pp.306-310

[23] Haiyan Zhao, Chuyi Song, Haili Zhao, Shizheng Zhang, "License plate recognition system based on morphology and LS-SVM," IEEE International Conference on , Granular Computing,26-28 Aug. 2008. GrC 2008., pp.826-829

[24] JianhongXie, "Optical Character Recognition Based on Least Squares Support Vector Machine", Third International Symposium on Intelligent Information TechnologyApplication, 21-22 November 2009, NanChang, China.

[25] Hong-Ying Yang, Xiang-Yang Wnag, Qin-Yan Wang, Xian-Jin Zhang, "LS-SVM based image segmentation using color and texture information", Elsevier Visual Image, Volume 23, 2012, pp.1095-1112.

[26] Johan A.K. Suykenes, Tony Van Gestel, Jos De Brabanter, Bart De Moor and JoosVandewalle, "Least Squares Support Vector Machine", World Scientic,2002.

[27] TaiwoOladipupoAyodele, "Types of Machine Learning Algorithms", New Advances in Machine Learning - Intech Publication,2010,pp.19-46.

[28] Cai-Xai Deng, Li-Xiang Xu and Shuali Li, "Classification of Support Vector machine and regression algorithm", New Advances in Machine Learning - Intech Publication, 2010,pp.75-91.

[29] TanveerAhsan, TaskeedJabid and Ui-Pil Chong, "Facial Expression Recognition Using Local Transitional Pattern on Gabor Filtered Facial Images",IETE Technical Review,30,1,2013,pp.47-52.

[30] J.G. Daugman,"Uncertainty relations for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters", Journal of the Optical Society of America,Vol. 2,1985, pp. 1160-1169.

[31]N. Petkov, P. Kruizinga, "Computational models of visual neurons specialized in the detection of periodic and aperiodic oriented visual stimuli: bar and grating cells", Biological Cybernetics, 76 (2), 1997, pp.83-96.

[32] N. Petkov,M.B. Wieling, "Gabor filter for image processing and computer vision" ,An Article.

[33] http://www.esat.kuleuven.ac.be/sista/lssvmlab/

[34] Seth Benton. Article on background subtraction methods www.eetimes.com, "Article on background subtraction methods"

[35] Available from : http://homepages.inf.ed.ac.uk/rbf/CAVIAR

[36] Background Modeling Available from: http://perception. i2ra- star.edu.sg/bk_model/ bkindex.html

On Annotation of Video Content for Multimedia Retrieval and Sharing

Mumtaz Khan*, Shah Khusro, Irfan Ullah
Department of Computer Science, University of Peshawar, Peshawar 25120, Pakistan

***Abstract*-The development of standards like MPEG-7, MPEG-21 and ID3 tags in MP3 have been recognized from quite some time. It is of great importance in adding descriptions to multimedia content for better organization and retrieval. However, these standards are only suitable for closed-world-multimedia-content where a lot of effort is put in the production stage. Video content on the Web, on the contrary, is of arbitrary nature captured and uploaded in a variety of formats with main aim of sharing quickly and with ease. The advent of Web 2.0 has resulted in the wide availability of different video-sharing applications such as YouTube which have made video as major content on the Web. These web applications not only allow users to browse and search multimedia content but also add comments and annotations that provide an opportunity to store the miscellaneous information and thought-provoking statements from users all over the world. However, these annotations have not been exploited to their fullest for the purpose of searching and retrieval. Video indexing, retrieval, ranking and recommendations will become more efficient by making these annotations machine-processable. Moreover, associating annotations with a specific region or temporal duration of a video will result in fast retrieval of required video scene. This paper investigates state-of-the-art desktop and Web-based-multimedia-annotation-systems focusing on their distinct characteristics, strengths and limitations. Different annotation frameworks, annotation models and multimedia ontologies are also evaluated.**

Keywords: Ontology, Annotation, Video sharing web application

I. INTRODUCTION

Multimedia content is the collection of different media objects including images, audio, video, animation and text. The importance of videos and other multimedia content is obvious from its usage on YouTube and on other platforms. According to statistics regarding video search and retrieval on YouTube [1], lengthy videos with average duration of 100 hours are uploaded to YouTube per minute, whereas 700 videos per day are shared from YouTube on Twitter, and videos comprising of length equal to 500 years are shared on Facebook from YouTube [1]. Figure1 further illustrates the importance and need of videos among different users from all over the world. As of 2014, about 187.9 million US Internet users watched approximately 46.6 million videos in March 2014 [2]. These videos are not only a good source of entertainment but also facilitate students, teachers, and research scholars in accessing educational and research videos from different webinars, seminars, conferences and encyclopedias. However, in finding relevant videos, the opinions of fellow users/viewers and their annotations in the form of tags, ratings, and comments are of great importance if dealt with carefully. To deal with this issue, a number of video annotation systems are available including YouTube, Vimeo[1], Youku[2], Myspace[3], VideoANT[4], SemTube[5], and Nicovideo[6] that not only allow users to annotate videos but also enable them to search videos using the attached annotations and share videos with other users of similar interests. These applications allow users to annotate not only the whole video, but also their specific event (temporal) and objects in a scene (pointing region). However, these are unable to annotate specific themes in a video, and browsing for specific scene, theme, event and object, and searching using whole video annotations are some of the daunting tasks that need further research.

One solution to these problems is to incorporate context-awareness using Semantic Web technologies including Resource Description Framework (RDF), Web Ontology Language (OWL), and several top-level as well as domain level ontologies in proper organizing, listing, browsing, searching, retrieving, ranking and recommending multimedia content on the Web. Therefore, the paper also focuses on the use of Semantic Web technologies in video annotation and video sharing applications. This paper investigates the state-of-the-art in video annotation research and development with the following objectives in mind:

- To critically and analytically review relevant literature regarding video annotation, video annotation models, and analyze the available video annotation systems in order to pin-point their strengths and limitations
- To investigate the effective use of Semantic Web technologies in video annotation and study different multimedia ontologies used in video annotation systems.
- To discover current trends in video annotation systems and place some recommendations that will open new research avenues in this line of research.

To the best of our knowledge this is the first ever attempt to the detailed critical and analytical investigation of the current trends in video annotation systems along with a detailed retrospective on annotation frameworks, annotation models and multimedia ontologies. Rest of the paper is organized as Section 2 presents state-of-the-art research and development in video annotation, video annotation frameworks and models, and different desktop- and Web-based video annotation systems. Section 3 contributes an evaluation framework for comparing the available video-annotation systems. Section 4 investigates different multimedia ontologies for annotating video content. Finally, Section 5 concludes our discussion and puts some recommendations before the researchers in this area.

---

[1] http://www.vimeo.com/
[2] http://www.youku.com/
[3] http://www.myspace.com/
[4] http://ant.umn.edu/
[5] http://metasound.dibet.univpm.it:8080/semtube/index.html
[6] http://www.nicovideo.jp/

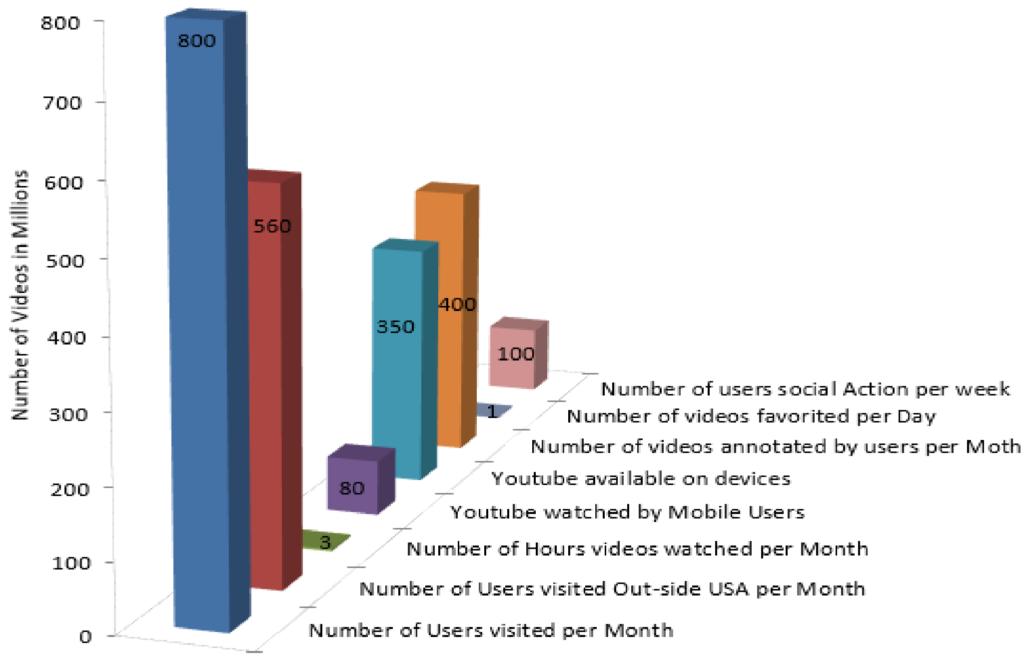


Figure 1. YouTube traffic and videos usage in the world.

II. STATE-OF-THE-ART IN VIDEO ANNOTATION RESEARCH

The idea behind annotation is not new, rather it has a long history of research and development and its origin can be traced back to 1945 when Vannevar Bush gave the idea of Memex that users can establish associative trails of interest by annotating microfilm frames [3]. In this section, we investigate annotations, video-annotations, video-annotation approaches and present state-of-the-art research and development practices in video-annotation systems.

A. *Using Annotations in Video Searching and Sharing*

Annotations are interpreted in different ways. They provide additional information in the form of comments, notes, remarks, expressions, and explanatory data attached to it or one of its selected parts that may not be necessarily in the minds of other users, who may happen to be looking at the same multimedia content [4, 5]. Annotations can be either formal (structured) form or informal (unstructured) form. Annotations can either be implicit i.e., they can only be interpreted and used by original annotators, or explicit, where they can be interpreted and used by other non-annotators as well. The functions and dimensions of annotations can be classified into writing vs reading annotations, intensive vs extensive annotations, and temporary vs permanent annotations. Similarly, annotations can be private, institutional, published, and individual or workgroup-based [6]. Annotations are also interpreted as universal and fundamental research practices that enable researchers to organize, share and communicate knowledge and collaborate with source material [7].

Annotations can be manual, semi-automatic or automatic [8, 9]. Manual annotations are the result of attaching knowledge structures that one bears in his mind in order to make the underlying concepts easier to interpret. Semi-automatic annotations require human intervention at some point while annotating the content by machines. Automatic annotations require no requiring human involvement or intervention. Table 1 summarizes annotation techniques with required levels of participation from humans and machines as well as gives some real world examples of tools that use these approaches.

TABLE 1 ANNOTATION APPROCHES

Annotation Techniques / Human Participation & Use of Tools	Manual	Semi-Automatic	Automatic
Human task	Entering some descriptive keywords	Entering initial query at start-up	No interaction
Machine task	Providing storage space or databases for storing and recording annotations	Parsing queries to extract information semantically to add annotations	Using recognition technologies for detecting labels and semantic keywords
Examples	GAT[7], Manatee[8], VIA[9], VAT[10]	Semantator[11], NCBO Annotator[12], cTAKES[13]	KIM[14], KAAS[15], GATE[16].

B. Annotation Frameworks and Models

Before discussing state-of-the-art video-annotation systems, it is necessary to investigate how these systems manage the annotation process by investigating different frameworks and models. These frameworks and models provide manageable procedures and standards for storing, organizing, processing and searching videos based on their annotations. These frameworks and models include Common Annotation Framework (CAF) [10], Annotea [11-14], Vannotea [15], LEMO [5, 14], YUMA [7], Annotation Ontology (AO) [16], Open Annotation Collaboration (OAC) [17], and Linked Media Framework (LMF) [18]. Common Annotation Framework (CAF), developed by Microsoft, annotates web resources in a flexible and standard way. It uses a web page annotation client named WebAnn that could be plugged into Microsoft Internet Explorer [10]. However, it can annotate only web documents and has no support for annotating other multimedia content.

Annotea is a collaborative annotation framework and annotation server that uses RDF database and HTTP front-end for storing annotations and responding to annotation queries. Xpointer is used for locating the annotations in the annotated document, Xlink is used for defining links between annotated documents whereas RDF is used for describing and interchanging metadata [11]. However, Annotea is limited as its protocol must be known to the client for accessing annotations, it does not take into account the different states of a web document, and has no room for annotating multimedia content. To overcome these shortcomings, several extensions have been developed. For example, Koivunnen [12] introduces additional types of annotations including bookmark annotations and topic annotations. Schroeter and Hunter [13] propose expressing multimedia content segments using content resources in

7http://upseek.upc.edu/gat/
8http://manatee.sourceforge.net/
9http://mklab.iti.gr/via/
10http://www.boemie.org/vat
11http://informatics.mayo.edu/CNTRO/index.php/Semantator
12http://bioportal.bioontology.org/annotator
13http://ctakes.apache.org/index.html
14http://www.ontotext.com/kim/semantic-annotation
15http://www.genome.jp/tools/kaas/
16http://gate.ac.uk/

connection with standard descriptions for establishing and representing the context such as Scalable Vector Graphics (SVG) or other MPEG-7 standard complex data types. Haslhofer et al [14] introduce annotation profiles that work as containers for content Annotea annotation type-specific extensions. They suggested that annotations should be de-referenceable resources on the Web by following Linked Data principles.

Vannotea is a tool for real-time collaborative indexing, description, browsing, annotation and discussion of video content [15]. It primarily focuses on providing support for real-time and synchronous video conferencing facilities and makes annotations simple and flexible to attain interoperability. This led to the adaptation of XML-based description schemes. It uses the Annotea, Jabber, Shibboleth and XACML technologies. W3C activity aims to advance the sharing of metadata on the Web. Annotea uses RDF and Xpointer for locating annotations within the annotated resource.

LEMO [5] is a multimedia annotation framework that is considered as the core model for several types of annotations. It also allows annotating embedded content items. This model uses MPEG-21 media fragment identification, but it supports only MPEG type media with rather complex and ambiguous syntactical structure when compared to W3C's media fragment URIs. Haslhofer et al [14] interlinked rich media annotations of LEMO to Linked Open Data (LOD) cloud. YUMA is another open web annotation framework based on LEMO that annotates whole digital object or its part and publishes annotation based on LOD principles [7].

Annotation Ontology (AO) [16] is an open annotation ontology developed in OWL and provides online annotations for web documents, images and their fragments. It is similar to OAC model but differs from OAC in terms of fragment annotations, representation of constraints as well as constraint targets as first-class resources. It also provides convenient ways for encoding and sharing annotations in FRD format. OAC is an open annotation model that annotates multimedia objects such as images, audio and video and allows sharing annotations among annotation clients, annotation repositories and web applications on the Web. Interoperability can be obtained by aligning this model with the specifications that are being developed within W3C Media Fragment URI group [17].

Linked Media Framework (LMF) [18] is concerned with how to publish, describe and interlink multimedia contents. The framework extends the basic principles of linked data for the publication of multimedia content and its metadata as linked data. It enables to store and retrieve contents and multimedia fragments in a unified manner. The basic idea of this framework is how to bring close together information and non-information resources on the basis of Linked Data, media management and enterprise knowledge management. The framework also supports annotation, metadata storage, indexing and searching. However, it lacks support for media fragments and media annotations. In addition, no attention has been given to rendering media annotations.

C. Video Annotation Systems

Today, a number of state-of-the-art video-annotation systems are in use, which can be categorized into desktop-based and Web-based video annotation systems. Here, we investigate these annotation systems with their annotation mechanisms, merits, and limitations.

1) Desktop-based Video-Annotation Systems:

Several desktop-based video annotation tools are in use allowing users to annotate multimedia content as well as to organize, index and search videos based on these annotations. Many of these desktop-based video-annotation tools use Semantic Web technologies including OWL and ontologies in properly organizing, indexing and searching videos based on annotations. These tools are investigated in the following paragraphs.

ANVIL [19, 20] is a desktop-based annotation tool that allows manually annotating multimedia content in linguistic research, gesture research, human computer interaction (HCI) and film studies. It provides descriptive, structural and administrative annotations for temporal segments, pointing regions or for entire source. The annotation procedure and XML schema specification are used to define the vocabulary. The head and body sections contain respectively the administrative and descriptive metadata having structural information for identifying temporal segments. Annotations are stored in XML format that can be easily exported to Excel and SPSS for statistical analysis. For speech transcription, data can also be imported from different phonetic tools including PRAAT and XWaves. ANVIL annotates only MPEG-1, MPEG-2, quick time, AVI, and MOV formats for multimedia content. However, the interface is very complex with no support for annotating specific objects and themes in a video. Searching for specific scene, event, object and theme of video is very difficult.

ELAN[17] [21] is a professional desktop-based audio and video annotation tool that allows users to create, edit, delete, visualize and search annotations for audio and video content. It has been developed specifically for language analysis, speech analysis, sign language and gestures or motions in audio/video content. Annotations are displayed together with their audio and/or video signals. Users create unlimited number of annotation tiers/layers. A tier is a logical group of annotations that places same constraints on structure, content and/or time alignment of characteristics. A tier can have a parent tier and child tier, which are hierarchically interconnected. It provides three media players namely Quick Time, Java Media Player, and Windows Media Player. In addition, it provides multiple timeline viewers to display annotations such as timeline viewer, interlinear viewer and grid viewer whereby each annotation is shown by a specific time interval. Its keyword search is based on regular expressions. Different import and export formats are supported namely shoebox/toolbox (.txt), transcriber (trs), chat (cha), preat (TextGrid), CSV/tab-delimited text (csv) and word list for the listing annotations. Some drawbacks include time-consumption because of manual annotation of videos, difficulty in use for users and multimedia content providers, complexity in interface, difficulty in learning for the ordinary users, lack of thematic-based annotations on video and difficulties in searching for specific scene, event, object or theme.

OntoELAN[18] [22] is an ontology-based linguistic multimedia annotation tool that inherits all the features of ELEN with some additional features. It can open and display ontologies in OWL language, and allows creating language profile for free-text and ontology-based annotations. OntoELAN has a time-consuming and complex interface. It does not provide thematic-based video annotations on videos, and searching for specific scene, event, object and theme in a video is difficult.

Semantic Multimedia Annotation Tool (SMAT) is a desktop-based video annotation tool used for different purposes including education, research, industry, and medical training. SMAT allows annotating MPEG-7 videos both automatically as well as manually and facilitates in arranging multimedia content, recognizing and tracing objects, configuring annotation sessions, and visualizing annotation and its statistics. However, because of complex

---

[17] http://tla.mpi.nl/tools/tla-tools/elan
[18] http://emeld.org/school/toolroom/software/software-detail.cfm?SOFTWAREID=480

interface, it is difficult to learn and use. User can annotate only those videos that are in flv and MPEG-7 format. Annotations are embedded in videos and therefore cannot be properly utilized. Searching for specific scene, event and theme of a video becomes difficult.

Video Annotation Tool (VAT) is a desktop-based application that allows manual annotation of MPEG-1 and MPEG-2 videos on frame by frame basis and in live recording. It allows users to import the defined OWL ontology files and annotates specific regions in a video. It also supports free text annotations on video, shot, frame by frame and on region level. However, it does not allow annotations on thematic and temporal basis. The interface is very complex and searching for a specific region, scene, event and theme is difficult.

Video and Image Annotation[19] (VIA) is a desktop-based application that allows manually annotating MPEG videos and images. It also allows for frame by frame video annotation during live recording. A whole video, shot, image or a specified region can be annotated using free-text or using OWL ontology. It uses video processing, image processing, audio processing, latent-semantic analysis, pattern recognition, and machine learning techniques. However, it has complex user interface and does not allow for temporal and thematic-level annotations. It is both time-consuming and resource-consuming whereas searching of a specific region, scene, event, and theme is difficult.

Semantic Video Annotation Suite[20] (SVAS) is desktop-based annotation tool that annotates MPEG-7 videos. SVAS combines features of media analyzer tool and annotation tool. Media analyzer is a pre-processing tool with automatic computational work of video analysis, content analysis, and metadata for video navigation where structure is generated for shot and key-frames and stored in MPEG-7 based database. The annotation tool allows users to edit structural metadata that is obtained from media analyzer and adds organizational and explanatory metadata on MPEG-7 basis. The organizational metadata describes title, creator, date, shooting and camera details of the video. Descriptive metadata contains information about persons, places, events and objects in the video, frame, segment or a region. It also enables users to annotate a specific region in a video/frame using different drawing tools including polygon and bounded box or deploying automatic image segmentation. This tool also facilitates automatic matching services for detecting similar objects in the video and a separate key-frame view is used for the results obtained. Thus users can easily identify and remove irrelevant key-frames in order to improve the retrieved results. It also enables users to copy annotations of a specific region to other region in a video that are same objects in a video by using a single click. This saves time as compared to time required by manual annotation. It exports all views in CVS format whereas MPEG-7 XML file is used to save the metadata [23].

Hierarchical Video Annotation System keeps video separate from its annotations and manually annotates AVI videos on scene, shot or frame level. It contains three modules including video control information module, annotation module and database module. The first module controls and retains information about video like play, pause, stop and replay. The second module is responsible for controlling annotations for which information regarding video is returned by first module. In order to annotate a video at some specific point, user needs to pause the video. At completion, the annotations are stored in the database. A Structured Query Description Language (SDQL) is also proposed which works like SQL but with some semantic reasoning. Although, the system enables

[19] http://mklab.iti.gr/via/
[20] http://www.joanneum.at/digital/produkte-loesungen/semantic-video-annotation.html

users to annotate specific portion of a video, it lacks support for object and theme-level annotations. Because of complex user interface, its usage is difficult and time-consuming. Similarly, searching for specific region and theme is difficult [24].

It can be easily concluded that these annotation systems are limited in a number of ways. Most of these systems have complex user interface with time-consuming and resource-consuming algorithms. We found no mechanisms for sharing the annotated videos on the Web. They cover only a limited number of video formats with no universal video annotator that supports almost any type of video format. . Most of these systems are limited in organizing, indexing, and searching videos on the basis of these annotations. There is almost no desktop-based system that can annotate a video on specific pointing-region, temporal duration and theme level. These systems also lack in using domain-level ontologies in semantic video annotation and searching.

2) Web-based Video-Annotation Systems:

A number of Web-based video annotation system are in use that allow users to access, search, browse, annotate, and upload videos on almost every aspect of life. For example, YouTube is one of the best and largest video-annotation systems where users upload, share, annotate and watch videos. The owner of the video can also annotate temporal fragments and specific objects in the video whereas other users are not allowed to annotate a video based on these aspects. These annotations establish no relationship between the annotations and specific fragments of the video. It expresses video fragments at the level of the HTML pages, which contain the video. Therefore, using YouTube temporal fragment, a user cannot point to the video fragment and is limited to the HTML page of that video [25, 26]. The system is also limited in searching specific object, event and theme in the video. Furthermore, the annotations are not properly organized and therefore, the owner of the video cannot find out flaws in the scene, event, object and theme.

VideoANT is a video annotation tool that facilitates students in annotating videos in flash format on temporal basis. It also provides the facilities of feedback of the annotated text to the users as well as to the multimedia content provider so that errors, if any, could be easily corrected [27]. However, VideoANT does not annotate videos on event, object and theme basis. Similarly, searching of specific event, object, and theme of a video is difficult.

EUROPEANA Connect Media Annotation Prototype (ECMAP[21]) [14] is an online annotation suite that uses Annotea to extend the existing bibliographic information about any multimedia content including audio, videos and images. It also facilitates cross multilingual search and cultural heritage at a single place. ECMAP supports free-text annotation of multimedia content using several drawing tools and allows for spatial- and temporal-based annotation of videos. Semantic tagging, enriching bibliographic information about multimedia content and interlinking different Web resources are facilitated through annotation process and LOD principles. Used in geo-referencing, ECMPA enables users in viewing high resolution maps, images and supports title-based fast delivery search. The target application uses YUMA[22] and OAC[23] models for implementing and providing these facilities to users. Some limitations include lacking support for adding thematic based annotation and searching for related theme.

---

[21] http://dme.ait.ac.at/annotation/
[22] http://yuma-js.github.com/
[23] http://www.openannotation.org/spec/beta/

Project Pad[24] is another collaborative video-annotation system containing set of tools for annotating, enhancing and enriching multimedia content for research, teaching and learning purposes. The goal is to support distance learning by making online notebook of the annotated media segments. It facilitates users in organizing, searching and browsing rich media content and makes teaching and learning easy by collecting the digital objects and making Web-based presentation of their selected parts, notes, descriptions and annotations. It supports scholarly reasoning, when the specific image is viewed or examined at the specific temporal period. It also supports synchronous interaction between teacher and students or among the small group of students. The students give the answer of questions and teacher examines their answers and records their progress. However, searching of specific theme, event and scene is difficult and there is no relationship among comments in videos.

SemTube is a video-annotation tool developed by SemLib project that aims at developing an MVC-based configurable annotation system pluggable with other web applications in order to attach meaningful metadata to digital objects [28, 29]. SemTube enhances the current state of digital libraries through the use of Semantic Web technologies and overcomes challenges in browsing and searching as well as provides interoperability and effective resource-linking using Linked Data principles. Videos are annotated using different drawing tools where annotations are based on fragment, temporal duration and pointing region. In addition, it provides collaborative annotation framework using RDF as a data model, media fragment URI and Xpointer, and is pluggable with other ontologies. However, SemTube has no support for theme-based annotations, and linking related scenes, events, themes and objects is difficult. Similarly, searching for specific event, scene, object and themes is not available.

Synote[25] [30, 31] multimedia annotation system publishes media fragments and user generated annotations using Linked Data principles. By publishing multimedia fragments and their annotations, Semantic Web agents and search engines can easily find these items. These annotations are also shared on social networks such as Twitter. It allows synchronized bookmarking, synmarks, comments, tags, notes with video and audio recordings whereas transcripts, slides and images are used to find and replay recording of video contents. While watching and listening to the lectures, transcripts and slides are displayed alongside. Browsing and searching for transcripts, synmarks, slide titles, notes and text content is also available. It stores annotation in XML format and facilitates users for public and private annotations. It uses media resources 1.0, Schema.org, Open Archives Initiative Object Reuse and Exchange (OAI-ORE), and Open Annotation Collaborative (OAC) in describing ontologies and in resource aggregation. However, it does not provide annotation on scene, event, object and theme levels in videos. The searching of specific scene, event and theme is difficult for the users.

KMI[26] is an LOD-based annotation tool developed by Department of Knowledge Media Institute[27] from Open University for annotating educational materials that come from different source in Open University. The resources include course forums, multi-participant audio/video environments for language and television programs on BBC. Users can annotate as well as search related information using LOD and other related technologies [32]. However,

---

[24] http://dewey.at.northwestern.edu/ppad2/index.html
[25] http://www.synote.org/synote/
[26] http://annomation.open.ac.uk/annomation/annotate
[27] http://www.kmi.open.ac.uk

the tool does not annotating theme, scene and object in video, and browsing of specific theme, object or scene is difficult.

Beside numerous advantages, Web-based video annotation systems are limited to exploit the use of annotations to the fullest. For example, using YouTube temporal fragment, a user cannot point to the video fragment and is limited to the HTML page of that video [25, 26]. There is no support for searching specific object, event and theme in the video. Due to improper organization of the annotations, the video uploader cannot find out flaws in the scene, event, object, and theme. These annotations establish no relationship between the annotations and specific fragments of the video. Similarly, VideoANT lacks in mechanisms for annotating objects and themes. ECMAP is limited in providing theme-based annotations with no support for searching related themes in a video. Project Pad is also limited in searching for specific theme, event, scene and object in a video and shows no relationships among comments in a video. Similarly, SemTube, Synote and KMI have no support for theme-based annotations, linking related scenes, events, themes and objects as well as for searching specific event, scene, theme and object. In order to further elaborate this discussion, we introduce an evaluation framework that compares different features and functions of these systems. The next section contributes this evaluation framework.

III. EVALUATING AND COMPARING VIDEO-ANNOTATION SYSTEMS

This section presents an evaluation framework consisting of different evaluation criteria and features, which compare the available video annotation systems and enable researchers to evaluate the existing video annotation systems. Table 2 presents different features and functions of video-annotation systems including annotation type, depiction, storage formats, target object type, vocabularies, flexibility, localization, granularity level, expressiveness, definition language, media fragment identification, and browsing and searching through annotation.

In order to simplify the framework, and make it more understandable, we give a unique number to a feature or function (Table 2) so that Table 3 can summarize the overall evaluation and comparison of these systems in a more manageable way. In Table 3, the first column enlists different video-annotation systems, whereas the topmost row contains features and functionalities of Table 2 as the evaluation criteria. The remaining rows contain the numbers that were assigned in Table 2 to different approaches, attributes and techniques used against the evaluation criteria. By closely looking at Table 3, we can easily conclude that none of the existing system supports advanced features that should be in any video-annotation systems. Examples of such features include searching for specific scene, object and theme using the attached annotations. Similarly, summarizing related scenes, objects and themes in video using annotations is not available. Furthermore, none of these systems has support for annotating specific theme in videos using either free-text or using ontology.

TABLE 2 FEATURES AND FUNCTIONALITIES OF VIDEO-ANNOTATION SYSTEMS

Features and Functions	Approaches/Techniques/Attributes
Annotation Depiction	(1) HTTP-derefrenceble RDF document, (2) Linked Data, (3) Linked Open Data, (4) embedded in content representation
Storage Formats	(5) XML, (6) Structure format (7) RDF, (8)MPEG-7/XML, (9) Custom XML, (10)OWL
Target Object Type	(11) web documents, (12) multimedia objects, (13) multimedia and web documents
Vocabularies used	(14) RDF/RDFS, (15) Media fragment URI, (16) OAC (Open annotation Collaborative), (17) Open Archives Initiative Object reuse and Exchange (OAI-ORE), (18) Schema.org, (19) LEMO, (20) FOAF (Friend of A Friend), (21) Dublin Core, (22) Timeline, (23) SKOS (simple knowledge organization system, (24) Free Text, (25) Keywords, (26) XML Schema, (27) Customized structural XML schema, (28) MPEG-7, (29) Cricket Ontology
Flexibility	(30) Yes, (31) No
Localization	(32)Time interval, (33) free hand, (34) pointing region
Granularity	(35) Video, (36) video segment, (37) frame, (38) moving region, (39) image, (40) still region, (41) event, (42) scene, (43) theme
Expressiveness	(44) Concept, (45) relations
Annotation Type	(46) Text, (47) Drawing tools, (48) public, (49) private
Definition languages	(50) RDF/RDFS, (51) OWL
Media Fragment Identification	(52) XPointer, (53) Media fragment URI 1.0, (54) MPEG-7 fragment URI, (55) MPEG-21 fragment URI, (56) Not Available
Browsing and Searching	(57) Specific scene, (58) Specific Object, (59) Theme, (60) Summaries related scenes, objects and Themes

TABLE 3 SUMMARIES AND ANALYSIS OF VIDEO-ANNOTATION SYSTEMS

Features		**Annotation depiction**	**Storage formats**	**Target Object Type**	**Vocabularies used**	**Flexibility**	**Localization**	**Granularity**	**Expressiveness**	**Annotation type**	**Definition languages**	**Media Fragment Identification**	**Browsing, Searching Scene, Event, Object, and Theme**	**Summarizing related videos based on Scene , Event, Object, and Theme**
Projects & Tools	YouTube	4	5,6,8	12	Nil	30	32,33,34	35,36,38, 39,41,42	Nil	46,47, 48, 49	Nil	56	Nil	Nil
	VideoANT	4	5,6	12	Nil	30	32,33	35,36,41, 42	Nil	46	Nil	56	Nil	Nil
	ANVIL	4	5	12	26,27	31	32	35,36	44	46	Nil	56	Nil	Nil
	SMAT	4	8	12	24,25, 28	31	32	35,36	44	46	Nil	56	Nil	Nil
	VAT	4	5,8	12	Nil	31	32,34	35,36,37, 38,41,42	44	46,47	51	56	Nil	Nil
	VIT	4	5,8	12	Nil	31	32,34	35,37,40,41, 42	44	46, 47	51	56	Nil	Nil
	OntoELAN	4	5	12	26,28	31	32	35	44	46	Nil	54	Nil	Nil
	Project Pad	4	5	12	24	31	32,34	35,36,39	Nil	46	Nil	56	Nil	Nil
	SemTube	2	5	12	14,15	30	32,34	35,36,37, 38,39,41, 42	44, 45	46,47	50	52,53	57	Nil
	Synote	4	5	11,1 2	15,16, 17,18	31	32	35,36,38	Nil	46	Nil	54	Nil	Nil
	ECMAP	3	5,7	13	16,19	30	32,33,34	35,36,38,39,41,42	44, 45	46,47, 48, 49	50	52,53, 54	57	Nil
	ELAN	4	9	12	24,25	31	32	35,36	44	46	Nil	56	Nil	Nil
	SVAS	4	8	12	24,25,28	31	32,34	35,36,37,39, 40,41	Nil	46,47	Nil	55	Nil	Nil
	KMI	3	5,7	12	20,21,22, 23	30	32,33,34	35,36,38, 39,41,42	44, 45	46	50,51	52,53	Nil	Nil
	Hierarchical Video Annotation System	4	6	12	Nil	31	32	35,36,37, 39,40, 41,42	44, 45	46	Nil	56	Nil	Nil

IV. USE OF MULTIMEDIA ONTOLOGIES IN VIDEO ANNOTATION

Although multimedia is being indexed, searched, managed and utilized on the Web, these activities are not much effective because the underlying semantics remain hidden. Multimedia needs to be semantically described for their easy and reliable discoverability and utilization by agents, web applications, and web services. Similarly, research and development have been very progressive in automatically segmenting and structuring multimedia content as well as recognizing their low-level features but producing multimedia data is problematic because it is complex and has a multidimensional nature. Metadata is used to represent the administrative, descriptive, and technical characteristics associated with multimedia objects.

There are many methods to describe multimedia content using metadata but these methods do not allow search across different repositories for a certain piece of content and they do not provide the facility to exchange content between different repositories. The metadata standard also increases the value of multimedia data that is used by different applications such as digital libraries, cultural heritage, education, and multimedia directory services etc. All these applications are used to share multimedia information based on semantics. For example, MPEG-7 standard is used to describe metadata about multimedia content. For this purpose, different semantics-based annotation systems have been developed that use Semantic Web technologies. Few of these are briefly discussed in the coming paragraphs.

ID3[28] is a metadata vocabulary for audio data and embedded in the MP3 audio file format. This contains title, artist, album, year, genre and other information about music files. It is supported by several software and hardware developers such as iTunes, Windows Media Player, Winamp, VLC and Pod, Creative Zen, Samsung Galaxy and Sony Walkman. The aim is to address a broad spectrum of metadata inlcuding the list of involved people, lyrics, band, and relative volume adjustment to ownership, artist, and recording dates. This metadata facilitates users in managing music files but service provider offers different tags. MPEG-7 is an international standard and multimedia content description framework used to describe different parts of multimedia content both low-level features and high-level features. It consists of different sets of description tools including Description Schemes (Dss), Descriptors (Ds), Description Definition Language (DDL). Description schemes describe audio/video features such as describing region, segments, object and events. Descriptors describe the syntax and semantics of audio and video contents, while DDL provides support for new descriptor and description schemes to be defined and existing description schemes to be modified [33-35]. MPEG-7 is also implemented using XML schemas and consists of 1182 elements, 417 properties, and 337 complex types [36]. However, this standard is limited because: (i) it is based on XML based schema so it is hard to be directly processed by the machine, (ii) it uses URNs which are cumbersome for the Web, (iii) it is not open standard to the Web such as RDF and OWL vocabulary, and (iv) it is debatable whether annotation by means of simple text string labels can be considered semantics [37].

MPEG-21[38, 39] is a suit of standards defining an open framework for multimedia delivery and consumption which supports a variety of businesses engaged in the trading of digital objects. It does not focus on the representation and coding of content like MPEG-1 to MPEG-7 but focuses on the filling of the gaps in the multimedia delivery chain. From metadata perspective, the parts of MPEG-21 describe rights and licensing of digital

[28] http://id3.org/

items. The vision of MPEG-21 is to enable transparent and augmented use of multimedia resources contained in digital items across a wide range of networks and devices. The standard is under development and currently contains 21 parts covering many aspects of declaring, identifying, and adapting digital items along the distribution chain including file formats and binary representation. The most relevant parts of MPEG-21 include part 2 known as Digital Item Declaration (DID) that provides an abstract model and an XML-based representation. The DID model defines digital items, containers, fragments or complete resources, assertions, statements, choices/selections, and annotations on digital items. Part 3 is Digital Item Identification and refers to complete or partial digital item descriptions. Part 5 is Right Expression Language (REL) that provides a machine-readable language to define rights and permissions using the concepts as defined in the Rights Data Dictionary. Part 17 is Fragment Identification for MPEG media types, which specifies syntax for identifying parts of MPEG resources on the Web. However, it specifies a normative syntax to be used in URIs for addressing parts of any resource but whose media type is restricted to MPEG.

As discussed, different metadata standards like ID3, and variants of MPEG are available however, the lack of better utilization of these metadata create difficulties including searching videos on the basis of specific theme, scene, and pointing regions. In order to semantically analyze multimedia data and make it searchable and reusable, ontologies are essential to express semantics in a formal machine process-able representation. Gruber [40] looks at ontology as "a formal, explicit specification of a shared conceptualization" [40]. Conceptualization refers to an abstract model of something and explicit means each element must be clearly defined and formal means the specification should be machine process-able. Basically ontologies are used to solve the problems that happen from using different nomenclature to refer to the same concepts, or using the same terms that refer to different concepts.

Ontologies are usually used to enrich and enhance browsing and searching on the Web. In addition, it is used to make different terms and resources on the Web meaningful to the information retrieval systems. It also helps in semantically annotating multimedia data on the Web. There are several upper-level and domain-specific ontologies that are particularly used in video-annotation systems. Figure 2 depicts a timeline of audio and video ontologies from 2001 to date, which have been developed and used in different video-annotation systems. These ontologies are freely available in the form of RDF(S) and OWL. These ontologies are discussed in the Section 4.1 and 4.2.

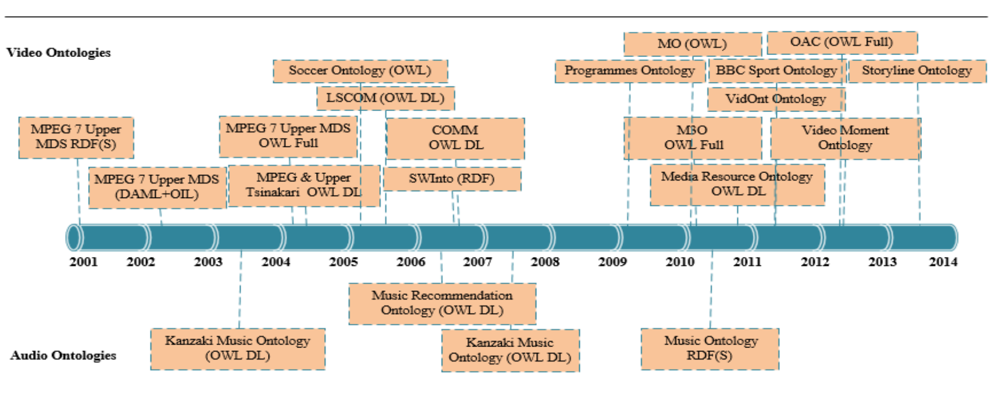


Figure 2: Timeline of audio and video ontologies

A. Upper-level Ontologies

A general-purpose or upper level ontology is used across multiple domains in order to describe and present general concepts [41]. Different upper-level ontologies have been developed and used for annotating multimedia content on the Web. However, some of the desktop-based video annotation systems also make use of different upper level ontologies. For example, MPEG-7 Upper MDS[29] [42] is an upper-level ontology available in OWL-Full that covers the upper part of the Multimedia Description Scheme (MDS) of the MPEG-7 standard comprising of 60 classes/concepts and 40 properties [42]. Hollink et al [43] added some extra features to the MPEG-7 Upper MDS for more accurate image analysis terms from the MATLAB image processing toolbox.

MPEG-7 Multimedia Description Scheme by Tasinaraki[30] [44] covers the full Multimedia Description Scheme (MDS) of the MPEG-7 standard. This ontology is defined in OWL DL and contains 240 classes and 175 properties. The interoperability of the complete MPEG-7 MDS is achieved with OWL, whereas the domain ontology of OWL can be integrated with MPEG-7 MDS for capturing the concepts of MPEG-7 MDS. The complex types in MPEG-7 correspond to OWL classes that represent groups of individuals sharing some properties. The complex types' attributes in MPEG-7 MDS are mapped to OWL data type properties whereas complex attributes are represented as OWL object properties that relate class instances. Relationships among OWL classes that correspond to complex types in MPEG-7 MDS are represented by instances of "RelationBaseTypes" class.

MPEG-7 Ontology [45] has been developed to manage multimedia data and to make MPEG-7 standard as a formal Semantic Web facility. MPEG-7 standard has no formal semantics and therefore, it was extended and translated to OWL. This ontology is fully automatically mapped from MPEG-7 standard to Semantic Web in order to give formal semantics. The aim of this ontology is to cover the whole standard and provide support for formal Semantic Web. Therefore, for this purpose, a generic mapping tool XSD2OWL has been developed. This tool converts the definitions of XML schema types and elements of the ISO standard into OWL definitions according to the set of rules given in [45]. This ontology consists of 525 classes, 814 properties, and 2552 axioms. This ontology can be used for upper-level multimedia metadata.

Large-Scale Concept Ontology for Multimedia[31] (LSCOM) [46] is a core ontology for broadcasting news video and contains more than 2500 vocabulary terms for annotating and retrieving broadcasted news video. This vocabulary contains information about objects, activities, events, scenes, locations, people, programs, and graphics. Under the LSCOM project, TREC conducted a series of workshops for evaluating video retrieval to encourage researchers regarding information retrieval by providing large test collections, uniform scoring techniques, and environment for comparing results. In 2012, various research organizations and researchers have completed one or more task regarding video content from different sources including semantic indexing (SIN), known-item search (KIS), instance search (INS), multimedia event detection (MED), multimedia event recounting (MER), and surveillance event detection (SED) [47].

Core Ontology for Multimedia (COMM) [48] is a generic and core ontology for multimedia contents based on MPEG-7 standard and DOLCE ontology. The development of COMM has changed the way of designing for

[29] http://metadata.net/mpeg7
[30] http://elikonas.ced.tuc.gr/ontologies/av_semantics.zip
[31] http://vocab.linkeddata.es/lscom/

multimedia ontologies. Before COMM, the efforts were focused on how to translate MPEG-7 standard to RDF/OWL. It was proposed to meet a number of requirements including compliance with MPEG-7 standard, semantic and syntactic interoperability, separation of concerns, and modularity and extensibility. COMM is developed using DOLCE and two ontology design patterns: Description and Situations (D&S), Ontology for Information (OIO). It has been implemented in OWL DL. The ontology facilitates users for multimedia annotations. The extended form of D&S and OIO patterns for multimedia data in COMM are consisted of decomposition pattern, media annotation patterns, content annotation patterns, and semantic annotation patterns. The decomposition patterns are used to describe the decomposition of video assets and handle the multimedia document structure whereas for annotating media, features, and semantic content of multimedia document, the media annotation pattern, content annotation pattern, and semantic annotation pattern are used [48].

Multimedia Metadata Ontology (M3O) [49, 50] describes and annotates complex multimedia resources. M3O is incorporated with different metadata standards and metadata models to provide semantic annotations with some additional development. M3O is meant to identify resources as well as to separate, annotate, and to decompose information objects and realizations. It is also aimed at representing provenance information. The aim of M3O is to represent data in rich form on the Web using Synchronized Multimedia Integration language (SMIL), Scalable Vector Graphics (SVG) and Flash. It fills the gap between structured metadata models and metadata standards such as XMP, JEITA, MPEG-7 and EXIF, and semantic annotations. It annotates the high-level and low-level features of media resources. M3O uses patterns that allow to accurately allocate arbitrary metadata to arbitrary media resources. M3O represents data structures in the form of Ontology Design Patterns that are based on formal upper-level ontology DOLCE[32] and DnS Ultralight (DUL). M3O reused the specialized patterns of DOLCE and DUL which are Description and Situation Pattern (D&S), Information and Realization Pattern, and Data Value Pattern. The main purpose of M3O is core ontology for semantic annotation but specially focused on media annotation. Furthermore, M3O consists of four patterns[33] (annotation, decomposition, collection, and provenance) that are respectively called Annotation Pattern, Decomposition Pattern, and Collection Pattern and Provenance Pattern. The annotations of M3O are represented in the form of RDF that can be embedded into SMIL for presentation of multimedia. The ontologies like COMM, Media Resource Ontology of the W3C and the image metadata standard exchangeable image file format (EXIF) are aligned with M3O. Currently the SemanticMM4U Framework uses M3O ontology as a general annotation model for multimedia data. It is also used for multi-channel generation of multimedia presentation in formats like SMIL SVG, Flash and others.

VidOnt [51] is a video ontology for making videos machine-processable and shareable on the Web. It also provides automatic description generation and can be used by filmmakers, video studios, education, e-commerce, and individual professionals. It is aligned with different ontologies like Dublin Core (DC), Friend of a Friend (FOAF) and Creative Commons (CC). VidOnt has been developed and implemented in different formats including RDF/XML, OWL, OWL functional syntax, Manchester syntax and Turtle.

---

[32] http://www.loa.istc.cnr.it/DOLCE.html
[33] http://m3o.semantic-multimedia.org/ontology/2010/02/28/annotation.owl

W3C Media Annotation Working Group[34] (MAWG) provides ontologies and APIs for facilitating cross-community data integration of information related to media objects on the Web. W3C has developed Ontology for Media Resources 1.0[35], which is a core vocabulary for multimedia objects on the Web. The purpose of this vocabulary is to join different descriptions about multimedia content and to give a core set of descriptive properties [52]. Its API[36] is available for the ontology that allows to access metadata stored in different formats and related to multimedia resources on the Web. The API is mapped with properties of metadata described in the ontology [53]. This API is both for the core ontology and for aligning the core ontology and metadata formats available on the Web for multimedia resources. The core ontology provides with interoperability for the applications used while using different multimedia metadata formats on the Web. The ontology contains different properties for a number of groups including identification, content description, keyword, rating, genre, distribution, rights, fragments, technical properties, title, locator, contributor, creator, date, and location. These properties describe multimedia resources available on the Web, which denote both abstract concepts such as “End of Watch” as well as specific instances in a video. However, it is not capable of distinguishing between different levels of abstraction that are available in some formats. The ontology when combined with the API ensures uniformity in accessing all its elements.

Smart Web Integrated Ontology[37] (SWInto) was developed within the Smart Web Project[38] from the perspective of open-domain question-answering and information seeking services on the Web. The ontology is defined in RDFS and based on a multiple layer partitioning into the partial ontologies. The main parts of the SWInto are based on DOLCE[39] and SUMO[40] ontologies. It contains different ontologies for knowledge representation and reasoning. It covers the following domain ontologies such as sports events that are represented in the sport event ontology, navigation, web services, discourse, linguistic information and multimedia data. SWInto integrates media ontology to represent multimodal information constructs [54].

Kanzaki[41] [55] is a music ontology designed and developed for describing classical music and visualizing performance. The ontology is composed of classes covering music, its instrumentations, events and related properties. The ontology distinguishes musical works from events of performance. It contains 112 classes, 34 properties and 30 individuals to represent different aspects of musical works and performances.

Music ontology [56, 57] describes audio related data such as artist, albums, tracks and the characteristics of the business related information about the music. This ontology is extended using existing ontologies including FRBR Final Report, Event Ontology, Timeline ontology, ABC ontology from the Harmony project and ontology from FOAF project. The Music Ontology can be divided into three levels of expressiveness. The first level supports information about tracks, artists, and release. The second level provides vocabulary about the music creation workflow such as composition, arrangement, performance, and recording etc. The third level provides information

---

[34] http://www.w3.org/2008/WebVideo/Annotations/
[35] http://www.w3.org/TR/mediaont-10/
[36] http://www.w3.org/TR/mediaont-api-1.0/
[37] http://smartweb.dfki.de/ontology_en.html
[38] http://www.smartweb-project.org/
[39] http://www.loa.istc.cnr.it/DOLCE.html
[40] http://www.ontologyportal.org/
[41] http://www.kanzaki.com/ns/music

about complex event decomposition such as particular performance in the event, and melody line of a particular work. It contains 141 classes, 260 object properties, 131 data type properties, and 86 individuals.

B. ***Domain-Specific Ontologies***

A domain-specific ontology describes a particular domain of interest in more detail and provides services to users of that domain. A number of domain-specific ontologies have been developed. For example, BBC has developed a number of domain-specific ontologies for different purposes. The News Storyline Ontology[42] is a generic ontology that is used to describe and organize news stories. It contains information about brands, series, episodes, and broadcasts etc. BBC Programme Ontology[43] defines concepts for programmes including brands, series or seasons, episodes, broadcast events, and broadcast services. The development of this ontology is based on the Music Ontology and FOAF Vocabulary. BBC Sport Ontology [58] is a simple and light weight ontology used to publish information about sports events, sport structure, awards and competitions.

Movie Ontology[44] (MO) is developed by Department of Informatics at the University of Zurich and licensed under the Creative Commons Attributions License. This ontology describes movies-related data including movie genre, director, actor and individuals. Defined in OWL, the ontology is further integrated to other ontologies that are provided in the Linked Data Cloud[45] to take advantage of collaboration effects [59-61]. This ontology is shared to LOD cloud and can be easily used by other applications and connected to other domains as well.

Soccer Ontology [62] consists of high level semantic data which has been developed to allow users to describe Soccer match videos and events of the game. This ontology is specially developed to be used with the IBM Video Annex Annotation tool to semantically annotate the soccer game videos. In addition, the ontology has been developed in the DDL and support MPEG related data. The Soccer Ontology was developed with a focus in facilitating the metadata annexing process.

Video Movement Ontology (VMO) describes any form of dance or human movement. This ontology is defined in OWL and using the Benesh Movement Notation (BMN) for ontology concepts and their relationships. The knowledge is embedded into ontology by using Semantic Web Rules Language (SWRL). The SWRL rules are used to perform rule-based reasoning on concepts. Additionally, we can search within VMO using SPARQL queries. It supports semantic description for image, sound and other objects [63].

V. CONCLUSION AND RECOMMENDATIONS

By reviewing the available literature and considering the state-of-the-art in multimedia annotation systems, we can conclude that the available features in these systems are limited and have not been fully utilized. Annotating videos based on themes, scenes, events and specific objects as well as their sharing, searching and retrieval is limited and need significant improvements. The effective and meaningful incorporation of semantics in these systems is still far away from their destiny.

---

[42] http://www.bbc.co.uk/ontologies/storyline
[43] http://www.bbc.co.uk/ontologies/po
[44] http://www.movieontology.org/
[45] http://linkeddata.org/

Current video-annotation systems have generated large amount of video content and annotations that are frequently accessed on the Web. Their frequent usage and popularity on the Web strongly recommend that multimedia content should be treated as first class citizen on the Web. However, the limitations in the available video-annotation web applications and the lack of sufficient semantics into their annotation process, organization, indexing and searching of the multimedia objects, make this dream far from realization. Users are not able to fully and easily use complex user interfaces of available systems and are unable to search for specific themes, events, or objects. Annotating specific object or annotating videos temporally or on pointing region basis is difficult to perform. This is also partly because of incompleteness in domain-level ontologies. Therefore, it becomes necessary to develop a video-annotation web application that allow users to annotate videos using free-text as well using a comprehensive and complete domain-level ontology.

In order to make this happen, we need to develop a collaborative video-annotation web application that allows its users to interact with multimedia content through a simple and user friendly user interface. Sufficient incorporation of semantics is required in order to convert annotations into semantic annotations. This way annotating specific object, event, scene, theme or whole video will become much easier. Users will be able to search for videos on a particular aspect or search within a video for particular theme, event, and object. Users will also be able to summarize related themes, objects, and events. Such an annotation system, if developed, can be easily extended to other domains and fields like YouTube that has different channels and categories of videos where user uploads videos to the concerned category or channel. Similarly, such a system can also be enriched and integrated with other domain-level ontologies.

By applying Linked Open Data principles we can annotate, search, and connect related scenes, objects and themes of videos that are available on different data sources. With the help of this technique, a lot of information will be linked with each other and video parts such as scenes, objects and themes could be easily approached and accessed. This will also enable users to expose, index, and search the annotated scenes, objects or themes and will be linked with global data sources.

REFERENCES

[1] YouTube, (2016) "Statistics about YouTube", available at: http://tuulimedia.com/68-social-media-statistics-you-must-see/ (accessed 22 February 2016).
[2] A. Lella, (2014) "comScore Releases March 2014 U.S. Online Video Rankings", available at: http://www.comscore.com/Insights/Press-Releases?keywords=&tag=YouTube&country=&publishdate=l1y&searchBtn=GO (accessed 25 December 2014).
[3] V. Bush's, (1945), "As We May Think. The Atlantic Monthly", available at: http://u-tx.net/ccritics/as-we-may-think.html#s8, (accessed 12 July 2014).
[4] K. O'Hara and A. Sellen, (1997), "A comparison of reading paper and on-line documents", in Proceedings of the ACM SIGCHI Conference on Human factors in computing systems, Los Angeles April 18-23, pp. 335-342.
[5] B. Haslhofer, W. Jochum, R. King, C. Sadilek, and K. Schellner, (2009), "The LEMO annotation framework: weaving multimedia annotations with the web", International Journal on Digital Libraries, vol. 10 No.1, pp. 15-32.
[6] C. C. Marshall, (2000), "The future of annotation in a digital (paper) world", in Proceedings of the 35th Annual Clinic on Library Applications of Data Processing: Successes and Failures of Digital Libraries, March 22-24, pp. 43-53.
[7] R. Simon, J. Jung, and B. Haslhofer, (2011),"The YUMA media annotation framework", in International Conference on Theory and Practice of Digital Libraries, TPDL, Berlin, Germany, September 26-28, 2011 , pp. 434-437.
[8] A. V. Potnurwar, (2012), "Survey on Different Techniques on Video Annotations", International Conference & workshop on Advanced Computing, June 2013.
[9] T. Slimani, (2013), "Semantic Annotation: The Mainstay of Semantic Web", arXiv preprint arXiv:1312.4794,
[10] D. Bargeron, A. Gupta, and A. B. Brush, (Access 2001), "A common annotation framework", Rapport technique MSR-TR-2001-108, Microsoft Research, Redmond, USA. Cité, vol. 3, p. 3.
[11] J. Kahan and M.-R. Koivunen, (Year), "Annotea: an open RDF infrastructure for shared Web annotations", in Proceedings of the 10th international conference on World Wide Web, 2001, pp. 623-632.
[12] M.-R. Koivunen, (2006), "Semantic authoring by tagging with annotea social bookmarks and topics", in Proc. of SAAW 2006, 1st Semantic Authoring and Annotation Workshop, Athens, Greece (November 2006).

[13] R. Schroeter, J. Hunter, and A. Newman, (2007),"Annotating relationships between multiple mixed-media digital objects by extending annotea", ed: Springer, pp. 533-548.
[14] B. Haslhofer, E. Momeni, M. Gay, and R. Simon, (2010), "Augmenting Europeana content with linked data resources", in Proceedings of the 6th International Conference on Semantic Systems, 2010, p. 40.
[15] R. Schroeter, J. Hunter, and D. Kosovic, (Year), "Vannotea: A collaborative video indexing, annotation and discussion system for broadband networks", in Knowledge capture, 2003, pp. 1-8.
[16] P. Ciccarese, M. Ocana, L. J. Garcia Castro, S. Das, and T. Clark, (2011), "An open annotation ontology for science on web 3.0", J Biomed Semantics, vol. 2, p. S4.
[17] B. Haslhofer, R. Simon, R. Sanderson, and H. Van de Sompel, (Year), "The open annotation collaboration (OAC) model", in Multimedia on the Web (MMWeb), 2011 Workshop on, 2011, pp. 5-9.
[18] T. Kurz, S. Schaffert, and T. Burger, (Year), "Lmf: A framework for linked media", in Multimedia on the Web (MMWeb), 2011 Workshop on, 2011, pp. 16-20.
[19] M. Kipp, (2001), "ANVIL - A Generic Annotation Tool for Multimodal Dialogue", Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech),
[20] M. Kipp, (2008), "Spatiotemporal coding in anvil", in Proceedings of the 6th international conference on Language Resources and Evaluation (LREC-08), 2008.
[21] H. Sloetjes, A. Russel, and A. Klassmann, (2007), "ELAN: a free and open-source multimedia annotation tool", in Proc. of INTERSPEECH, 2007, pp. 4015-4016.
[22] A. Chebotko, Y. Deng, S. Lu, F. Fotouhi, A. Aristar, H. Brugman, A. Klassmann, H. Sloetjes, A. Russel, and P. Wittenburg, (2004), "OntoELAN: An ontology-based linguistic multimedia annotator", in Multimedia Software Engineering, 2004. Proceedings. IEEE Sixth International Symposium on, 2004, pp. 329-336.
[23] P. Schallauer, S. Ober, and H. Neuschmied, (2008), "Efficient semantic video annotation by object and shot re-detection", in Posters and Demos Session, 2nd International Conference on Semantic and Digital Media Technologies (SAMT). Koblenz, Germany, 2008.
[24] C. Z. Xu, H. C. Wu, B. Xiao, P.-Y. Sheu, and S.-C. Chen, (2009), "A hierarchical video annotation system", in Semantic Computing, 2009. ICSC'09. IEEE International Conference on, 2009, pp. 670-675.
[25] W. Van Lancker, D. Van Deursen, E. Mannens, and R. Van de Walle, (2012), "Implementation strategies for efficient media fragment retrieval", Multimedia Tools and Applications, vol. 57, pp. 243-267.
[26] YouTube, "YouTube Video Annotations", available at: http://www.youtube.com/t/annotations_about (7 January 2014)
[27] B. Hosack, (2010), "VideoANT: Extending online video annotation beyond content delivery", Tech Trends, vol. 54, pp. 45-49.
[28] M. Grassi, C. Morbidoni, and M. Nucci, (2011),"Semantic web techniques application for video fragment annotation and management", ed: Springer, pp. 95-103.
[29] C. Morbidoni, M. Grassi, M. Nucci, S. Fonda, and G. Ledda, (2011), "Introducing SemLib project: semantic web tools for digital libraries", in International Workshop on Semantic Digital Archives-Sustainable Long-term Curation Perspectives of Cultural Heritage Held as Part of the 15th International Conference on Theory and Practice of Digital Libraries (TPDL), Berlin, 2011.
[30] Y. Li, M. Wald, G. Wills, S. Khoja, D. Millard, J. Kajaba, P. Singh, and L. Gilbert, (2011), "Synote: development of a Web-based tool for synchronized annotations", New Review of Hypermedia and Multimedia, vol. 17, pp. 295-312.
[31] Y. Li, M. Wald, T. Omitola, N. Shadbolt, and G. Wills, (2012), "Synote: weaving media fragments and linked data",
[32] D. Lambert and H. Q. Yu, (2010), "Linked Data based video annotation and browsing for distance learning",
[33] S.-F. Chang, T. Sikora, and A. Purl, (2001), "Overview of the MPEG-7 standard", Circuits and Systems for Video Technology, IEEE Transactions on, vol. 11, pp. 688-695.
[34] J. M. Martinez, (2004) "MPEG-7 overview (version 10), ISO", available at: (Accessed
[35] P. Salembier, T. Sikora, and B. Manjunath, (2002), Book Introduction to MPEG-7: multimedia content description interface,John Wiley & Sons, Inc., Place of John Wiley & Sons, Inc.
[36] R. Troncy, B. Huet, and S. Schenk, (2011), Book Multimedia semantics: metadata, analysis and interaction,John Wiley & Sons, Place of John Wiley & Sons.
[37] R. Arndt, R. Troncy, S. Staab, L. Hardman, and M. Vacura, (2007), Book COMM: designing a well-founded multimedia ontology for the web,Springer, Place of Springer.
[38] J. Bormans and K. Hill, (2002), "MPEG-21 Overview v. 5", ISO/IEC JTC1/SC29/WG11 N, vol. 5231, p. 2002.
[39] I. Burnett, R. Van de Walle, K. Hill, J. Bormans, and F. Pereira, (2003), "MPEG-21: goals and achievements",
[40] T. R. Gruber, (1993), "A translation approach to portable ontology specifications", Knowledge acquisition, vol. 5, pp. 199-220.
[41] T. Sjekavica, I. Obradović, and G. Gledec, (2013), "Ontologies for Multimedia Annotation: An overview", in 4th European Conference of Computer Science (ECCS'13), 2013.
[42] J. Hunter, (2005), Book Adding multimedia to the Semantic Web-Building and applying an MPEG-7 ontology,Wiley, Place of Wiley.
[43] L. Hollink, S. Little, and J. Hunter, (2005), "Evaluating the application of semantic inferencing rules to image annotation", in Proceedings of the 3rd international conference on Knowledge capture, 2005, pp. 91-98.
[44] C. Tsinaraki, P. Polydoros, and S. Christodoulakis, (2004),"Interoperability support for ontology-based video retrieval applications", ed: Springer, pp. 582-591.
[45] R. García and Ò. Celma, (2005), "Semantic integration and retrieval of multimedia metadata", in 2nd European Workshop on the Integration of Knowledge, Semantic and Digital Media, 2005, pp. 69-80.
[46] M. Naphade, J. R. Smith, J. Tesic, S.-F. Chang, W. Hsu, L. Kennedy, A. Hauptmann, and J. Curtis, (2006), "Large-scale concept ontology for multimedia", Multimedia, IEEE, vol. 13, pp. 86-91.
[47] P. Over, G. Awad, J. Fiscus, B. Antonishek, M. Michel, A. F. Smeaton, W. Kraaij, and G. Quénot, (2011), "TRECVID 2011-an overview of the goals, tasks, data, evaluation mechanisms and metrics", in TRECVID 2011-TREC Video Retrieval Evaluation Online, 2011.
[48] R. Arndt, R. Troncy, S. Staab, and L. Hardman, (2009),"COMM: A Core Ontology for MultimediaAnnotation", ed: Springer, pp. 403-421.
[49] C. Saathoff and A. Scherp, (2009), "M3O: the multimedia metadata ontology", in Proceedings of the workshop on semantic multimedia database technologies (SeMuDaTe), 2009.
[50] C. Saathoff and A. Scherp, (2010), "Unlocking the semantics of multimedia presentations in the web with the multimedia metadata ontology", in Proceedings of the 19th international conference on World wide web, 2010, pp. 831-840.
[51] L. F. Sikosa, (2011), "VidOnt: the video ontology",

[52] W. Lee, W. Bailer, T. Bürger, P. Champin, V. Malaisé, T. Michel, F. Sasaki, J. Söderberg, F. Stegmaier, and J. Strassner, (2012), "Ontology for media resources 1.0. W3C recommendation", World Wide Web Consortium,
[53] F. Stegmaier, W. Lee, C. Poppe, and W. Bailer, "API for media resources 1.0. W3C Working Draft," ed, 2011.
[54] D. Oberle, A. Ankolekar, P. Hitzler, P. Cimiano, M. Sintek, M. Kiesel, B. Mougouie, S. Baumann, S. Vembu, and M. Romanelli, (2007), "DOLCE ergo SUMO: On foundational and domain models in the SmartWeb Integrated Ontology (SWIntO)", Web Semantics: Science, Services and Agents on the World Wide Web, vol. 5, pp. 156-174.
[55] M. Kanzaki, "Music Vocabulary," ed, 2007.
[56] F. Giasson and Y. Raimond, (2007), "Music ontology specification", Specification, Zitgist LLC,
[57] Y. Raimond, S. Abdallah, M. Sandler, and F. Giasson, (2007), "The music ontology", in Proceedings of the International Conference on Music Information Retrieval, 2007, pp. 417-422.
[58] P. W. Jem Rayfield, Silver Oliver, (2012) "Sport Ontology", available at: http://www.bbc.co.uk/ontologies/sport/ (10-11-2012)
[59] A. Bouza, (2010), "MO–the Movie Ontology", MO–the Movie Ontology,
[60] R. M. Sunny Mitra, Suman Basak, Sobhit Gupta, "Movie Ontology", available at: http://intinno.iitkgp.ernet.in/courses/1170/wfiles/243055.
[61] S. Avancha, S. Kallurkar, and T. Kamdar, (2001), "Design of Ontology for The Internet Movie Database (IMDb)", Semester Project, CMSC, vol. 771,
[62] L. A. V. da Costa Ferreira and L. C. Martini, "Soccer ontology: a tool for semantic annotation of videos applying the ontology concept",
[63] D. De Beul, S. Mahmoudi, and P. Manneback, (2012), "An Ontology for video human movement representation based on Benesh notation", in International Conference on Multimedia Computing and Systems (ICMCS), 2012, pp. 77-82.

A New Approach for Energy Efficient Linear Cluster Handling Protocol In WSN

Jaspinder Kaur
M.Tech Student

Varsha Sahni
Assistant professor

Abstract

*Wireless Sensor Networks (WSN) is a rising field for researchers in the recent years. For obtaining durability of network lifetime, and reducing energy consumption, energy efficiency routing protocol play an important role. In this paper, we present an innovative and energy efficient routing protocol. A New linear cluster handling (LCH) technique towards Energy Efficiency in Linear WSNs with multiple static sinks [4] in a linearly enhanced field of 1500m*350m^2. We are divided the whole into four equal sub-regions. For efficient data gathering, we place three static sinks i.e. one at the centre and two at the both corners of the field. A reactive and Distance plus energy dependent clustering protocol Threshold Sensitive Energy efficient with Linear Cluster Handling [4] DE (TEEN-LCH) is implemented in the network field. Simulation shows improved results for our proposed protocol as compared to TEEN-LCH, in term of throughput, packet delivery ratio and energy consumption.*

Keywords: *WSN; Routing Protocol; Throughput; Energy Consumption; Packet Delivery Ratio*

1. Introduction

Wireless Sensor Networks (WSNs) are collection of small sensors with restricted energy resources. Based on the cheap cost of these devices, it is easier to deploy a large number of nodes to monitor a large area. Size and cost limitations on resources such as memory,

Applications of WSNs include military surveillance, industries control and monitoring sensing, traffic control, wildfire observation, etc. earlier routing techniques, like Minimum Transmission Energy (MTE) and Direct Communication (DC), as the recent cluster based techniques these are not as energy-efficient, because every node forward its sensed information directly to the Base Station (BS) in the DC. Nodes far from the BS die out more quickly and network lifetime is reduced. MTE is better from DC because node communication with its nearest neighbor. Furthermore, long distance nodes are prohibiting by DC, while in MTE closer nodes battery power drains out more quickly. Cluster based routing protocol; (LEACH) was proposed, to overcome all these deficiencies, which minimized the energy consumption.

Lifetime of WSNs and Efficiency based upon the design of the protocol. Packet Delivery Ratio, Energy Consumption, Throughput of the whole network is much upgrade in recent techniques. In recent era, nearly all techniques chase routing protocols which is based on clustering. All nodes are located in a cluster and it receives data in their cluster, selection of CH is done on the different bases. In the cluster it receives data of all nodes and transfer it towards the base station in the form of data packets, from station end user can access it easily. To decrease the use of energy, data aggregation is executed by the CH. In this technique, extra data packets are sending to BS and network lifetime is enhanced.

In this paper, we establish a multi-sink protocol in a linearly enhanced field. We divided the network area into equal regions and equal number of nodes having similar quantity of energy is deployed in each region which makes network homogenous. As in our proposed protocol, multi-sink are used therefore, CHs in every region forward their data to nearby static sinks. Hence, division of network area into multiple regions and multiple static sinks approach enhance the remaining energy and throughput of the network.

The rest of the paper are arranged as follows: in section [4] 2 the related work is explained. Section 3 discusses the motivation and points out the derivation of the recent works. The section 4 gives the detail of the proposed work. In section 5, simulations are discussed and different parameters are analyzed with their plots. Then section 6 is about the conclusion of proposed technique.

2. RELATED WORK

Authors in this paper [1] focus on mainly driven over the survey of the hierarchical cluster-based available routings in Wireless Sensor Network for energy consumption. Low-Energy Adaptive Clustering Hierarchy (LEACH) protocol is one of the best hierarchical protocols utilizing the probabilistic model to manage the energy consumption of WSN. Simulation results, shows themenergy consumption over time of three nodes with distance to the Base Station.

In [2] authors studied the Routing protocol of wireless sensor network research is the key problem, according to network topology, routing protocols can be divided into flat and hierarchical routing protocol. From the basic ideas, the advantages and disadvantages and applications the article introduces several typical hierarchical routing protocols in detail,

In this work, [3] authors propose Quadrature-LEACH (Q LEACH) for homogenous networks which enhances stability period, network life-time and throughput quiet significantly.

In this paper [4], authors present a scalable and energy efficient routing protocol, A New Linear Cluster Handling (LCH) Technique Towards energy efficiency. In linear WSNs with multiple static sinks in linearly enhanced field of 1000m*2m. The whole network field is divided into four equal sub-regions. For efficient data gathering, place three static sinks i.e. two at the both corners and one at the centre of the field. A proactive routing protocol Distributed Energy Efficient Clustering with Linear Cluster Handling (DEEC-LCH) is implemented in the network field. Furthermore, a reactive protocol Threshold Sensitive Energy Efficient with Linear Cluster Handling (TEEN-LCH) is also implemented for the same scenario with three static sinks . Simulation show s improved results for our proposed protocols as compared to simple DEEC and TEEN, in term of network lifetime, Throughput and energy consumption

In this paper, [5] authors propose a General Self-Organized Tree-Based Energy-Balance routing protocol (GSTEB) which builds a routing tree using a process where, for each round, BS assigns a root node and broadcasts this selection to all sensed nodes. Subsequently, each node selects its parent by considering only itself and its neighbors' information, thus making GSTEB a dynamic protocol. Simulation results show that GSTEB has a better performance than other protocols in balancing energy consumption, thus prolonging the lifetime of WSN.

This paper [6] elaborately compares two important clustering protocols, namely LEACH and LEACH-C (centralized), using NS2 tool for several chosen scenarios, and analysis of simulation results against chosen performance metrics with latency and network lifetime being major among them. The paper will be concluded by mentioning the observations made from analyses of results about these protocols.

In this paper, [7] authors propose a new technique for the selection of the sensors cluster-heads based on the amount of energy remaining after each round [(4), (5)]. As the minimum percentage of energy for the selected leader is determined in advance and consequently limiting its performance and nonstop coordination task, the new hierarchical routing protocol is based on an energy limit value threshold preventing the creation of a group leader, to ensure reliable performance of the whole network.

International Journal of xxxxxx
Vol. x, No. x, (20xx)

3. MOTIVATION

In order to increase the lifetime of network, there are two possible approaches. The first approach is to expand the energy of sensor nodes that becomes the device more expensive. The second option is to decrease the quantity of data, but this would reduce the throughput. The most similar routing protocol as compared to DC and MTE, The LEACH devices a novel way to expand the network lifetime and throughput. However, LEACH is not executed in linearly enhanced network. At the centre of field Single static sink reduce the lifetime and throughput of the network. In our proposed protocol, with multiple static sinks we divide the field into equal regions in the network and the cluster head chosen on the basis of energy and distance of every node in each sub-region which enhanced the network efficiency.

4. THE PROPOSED PROTOCOL

In this section, we present our proposed protocol DE (TEEN-LCH) in which cluster head is chosen on the bases of Energy and Distance of the nodes in the network. Description of DE (TEEN-LCH) is given in the following subsections.

No. of Item description specification	No. of Item description specification
Simulation Area	$1500m*350m^2$
No. of nodes	47
Channel type	Channel/Wireless Channel
Simulation time	35.0sec
Antenna model	Antenna/Omni Antenna
Link Layer Type	LL
Energy Model	Energy Model
MAC type	Mac/802_11
Interface queue type	Queue/ Drop Tail /Pri Queue
radio-propagation model	Propagation/Two Ray Ground
network interface type	Phy/Wireless Phy

A. Region Formation

The multiple sinks are using in order to produce proper transmission; proposed protocol is linearly elevated in the field of $1500m*350m^2$. Deployed equal sized sub-regions and split whole network area into these sub-regions. In this each sub-region, an autonomic cluster is created which eventually decrease the transmission distance as well as energy consumption.

B. Deployment of nodes and sinks positions

The important task after sub-region formation is to deploy nodes in the field in such a way that maximum area can be surrounded by the nodes. Equal numbers of nodes are deployed in each sub-region. Three sinks are placed in the network, two at the both corners of the field and one at the centre of the field. In this way, CH receives sensed data of the nodes and sends it to its nearby sink in the field.

https://dx.doi.org/10.6084/m9.figshare.3153940 https://sites.google.com/site/ijcsis/
ISSN 1947-5500

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 3, March 2016

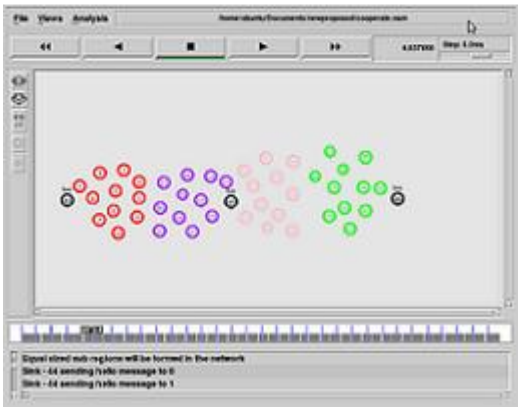
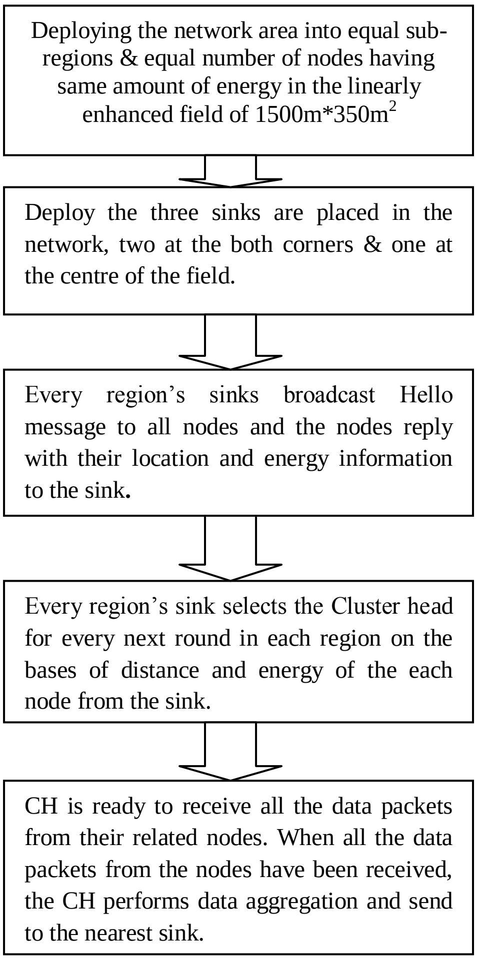


Fig 1: Region formation and Sink placement.

C. Protocol Operation



The protocol working is divided into separate phases such as;

- Advertisement Phase
- Cluster setup phase

- Data transmission phase

D. Advertisement phase

The proposed protocol earns credit by implementation multi-sinks and cluster formation in each region, which result in extension of throughput and lifetime of the network. Every region's sinks broadcast Hello message to all nodes and the nodes reply with their location and energy information to the sink. The sink selects the cluster head for every next round by using this information.

E. Cluster setup phase

Initially, when clusters are formed in a region, each node sends their location and energy information in the reply of sink's hello broadcast message, and then the sink select the node as a cluster head whose energy is more and the distance is less from the sink.

F. Data Transmission

Once sub-regions are formed and clusters are selected, then data transmission is started. CH is ready to receive all the data packets from their related nodes. When all the data packets from the nodes have been received, the CH performs data aggregation and send to the sink. As the Base Station is nearby every sub-region, so it requires low transmission energy. Same procedure is executed in every sub-region.

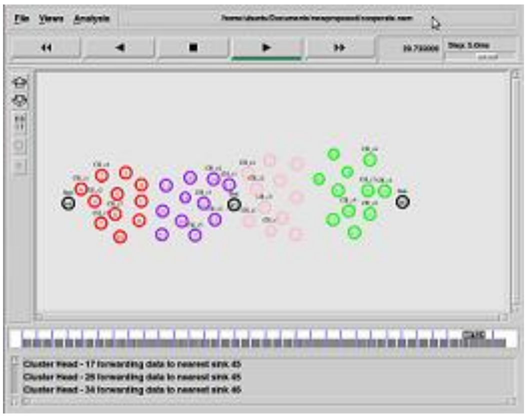


Fig 2: Cluster Setup and data transmission.

5. SIMULATION RESULTS

Performance of proposed protocol DE (TEEN-LCH) is representing on the basis of different parameters. Whole region of $1500m*350m^2$ is divided into four sub-regions in which equal number of nodes is randomly deployed. Three sinks are placed in the network at different locations and Cluster Head of every sub-region sends data packets to its closest sink respectively.

A. Packet Delivery ratio

It is defined as the ratio of no. of packets received to the no. of packets sent in the network to the base station. The greater value of the packet delivery ratio means better performance of protocol. The new proposed protocol DE (TEEN-LCH) has the greater value of the packet delivery ratio which is prove that the better performance in comparison of TEEN-LCH. This better performance is because of advertisement phase is performed only in first round by the sink in each sub-region and choose the cluster head of every round. So the congestion is decreased in the network and the packet delivery ratio is increased.

PDR= received_packets / generated_packets

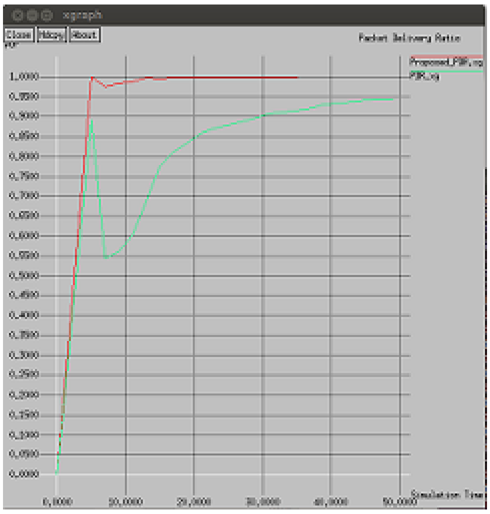


Fig 3: Packet Delivery ratio

TABLE I.

PROTOCOL	TIME	PDR
TEEN-LCH	49sec	94%
DE(TEEN-LCH)	35SEC	99%

B. Energy Consumption

The energy consumption is the aggregate of used energy by all the nodes in the network, where the used energy of a node is the sum of the energy used for transmission, including sending, receiving, and idling. In comparison of TEEN-LCH and DE (TEEN-LCH), the remaining energy of the new proposed protocol DE (TEEN-LCH) is more because of cluster head selection is performed on the basis of ratio of energy and distance of each node from the sink in each sub-region. So in each sub-region sink broadcast hello message for advertisement only in first round to each node and nodes reply with their energy and distance information to the sink, and sink choose the Cluster Head of every round in first round and the energy consumption is decreased. The more remaining energy provides the more stability period and network lifetime.

Energy {exp $ initial energy ($i)-$ final energy ($f)}

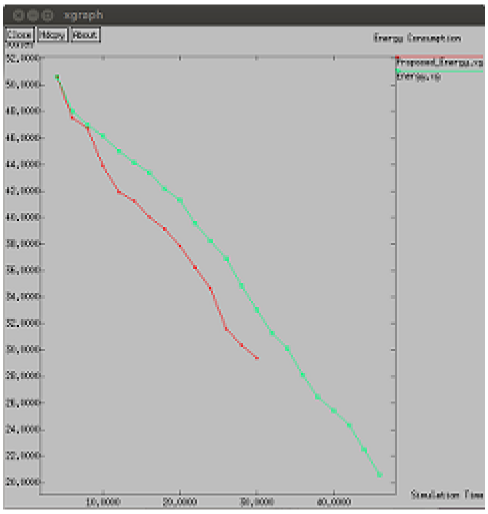


Fig 4: Energy Consumption

TABLE II.

PROTOCOL	TIME	REMAINING ENERGY
TEEN-LCH	46sec	20.5992JOULES
DE(TEEN-LCH)	35SEC	31.5012JOULES

C. Throughput

Throughput is the average of data packets received at the destination (i.e. at base station). The new proposed protocol DE (TEEN-LCH) shows the improved throughput value as compared to TEEN-LCH. In new proposed protocol the network has greater value of the packet delivery ratio because of the advertisement phase is done only in first round rather than in every round by the sink in each sub-region. The more packet delivery ratio provides the improved throughput value in the network.

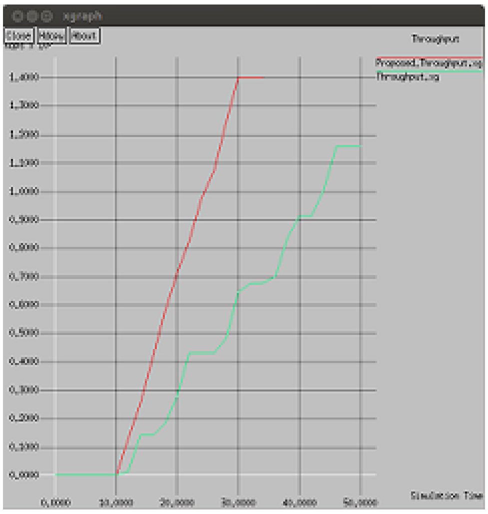


Fig 5: Throughput.

Throughput=received data*8 / data transmission period

TABLE III.

PROTOCOL	TIME	THROUGHPUT
TEEN-LCH	50sec	1159.17KBPS
DE(TEEN-LCH)	34SEC	1400.83KBPS

6. CONCLUSION

We proposed DE (TEEN-LCH) an energy-aware adaptive multi-sink routing protocol used in linearly enhanced field. In each region equal numbers of nodes are randomly deployed. Three sink are placed on the three different places in the network these sink receive data packets from their nearest nodes and CHs. CH is selected by the each region's sink in every region for each round with the help of advertisement phase which receives sensed data of nodes and after aggregation transfer it to Base Station. In the same way, results present that the proposed strategy increases the packet delivery ratio and improves the throughput. In future, we are interested to implement mobile sinks with chain based routing.

7. AKNOWLEDGMEN

The authors wish to thank the faculty from the computer science department at CTIEMT, Jalandhar for their continued support and feedback.

REFERENCES

[1] M.Shanthi Dr. E.RamaDevi "A Cluster Based Routing Protocol in Wireless Sensor Network for Energy Consumption" Int. J. Advanced Networking and Applications Volume: 05, Issue: 04, Pages:2015-2020 (2014) ISSN : 0975-0290 .

[2] DaWei Xu, Jing Gao,a "Comparison Study to Hierarchical Routing Protocols in Wireless Sensor Networks" ELSEVIER Procedia Environmental Sciences 10 (2011) 595 – 600.

[3] B. Manzoor, N. Javaid, O. Rehman, M. Akbar, Q. Nadeem, A. Iqbal, M. Ishfaq" Q-LEACH: A New Routing Protocol for WSNs " Procedia Computer Science 19 (2013) 926 – 931

[4] M. Sajid, K. Khan, U. Qasim, Z.A. Khaan,S. Tariq, N. Javaid,"A New Linear Cluster Hnadling(LCH)Technique Toward's Energy efficiency in Linear WSNs" 2015 IEEE 29 International Conference on Advanced Information Networking and Applications.

[5] Zhao Han, Jie Wu, *Member*, *IEEE*, Jie Zhang, Liefeng Liu, and Kaiyun Tian, "A General Self-Organized Tree-BasedEnergy-Balance Routing Protocol" IEEE TRANSACTIONS ON NUCLEAR SCIENCE, VOL. 61, NO. 2, APRIL 2014.

[6] Geetha. V., Pranesh.V. Kallapur, Sushma Tellajeera," Clustering in Wireless Sensor Networks: Performance Comparison of LEACH & LEACH-C Protocols Using NS2" ELSEVIER Procedia Technology 4 (2012) 163 – 170.

[7] Salim EL KHEDIRIa,b,c, Nejah NASRIa, Anne WEIc, Abdennaceur KACHOURId," A New Approach for Clustering in Wireless Sensors Networks Based on LEACH" ELSEVIER Procedia Computer Science 32 (2014) 1180 – 1185.

[8] Jaspinder Kaur, Varsha Sahni "Survey on Hierarchical Cluster Routing Protocols of WSN" International Journal of Computer Applications (0975 – 8887) Volume 130 – No.17, November2015.

[9] F. Xiangning, S. Yulin. "Improvement on LEACH Protocol of Wireless Sensor Network", 2007, International Conference on Sensor Technologies and Applications pp 260-264, ido:10.1109/SENSORCOMM.2007.60 .

[10] Kemal Akkaya , Mohamed Younis "A survey on routing protocols for wireless sensor networks" ELSEVIER.

[11] Jaspinder Kaur, Taranvir Kaur, Kanchan Kaushal "Survey on WSN Routing Protocols" International Journal of Computer Applications (0975 – 8887) Volume 109 – No. 10, January 2015 .

[12] Varsha Sahni ,Jaspreet Kaur, Sonia Sharma "Security Issues and Solutions in Wireless Sensor Networks" International Journal of Computer Science and Information Technologies, Vol. 3 (2) , 2012.

[13] Nikolaos A. Pantazis, Stefanos A. Nikolidakis and Dimitrios D.Vergados, Senior Member, IEEE" Energy-Efficient Routing Protocols in Wireless Sensor Networks: A Survey" IEEE communications survey & tutorials, vol. 15, no. 2, second quarter 2013.

[14] Harjeet Kaur 1, Manju Bala 2 ,Varsha Sahni 3 "Study of Black hole Attack Using Different Routing Protocols in MANET" International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 2, Issue 7, July 2013.

[15] Sanjeev Kumar Gupta, Neeraj Jain, Poonam Sinha " Clustering Protocols in Wireless Sensor Networks: A Survey" International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 5– No.2, January 2013.

[16] S. Misra et al. (eds.), Guide to Wireless Sensor Networks, Computer Communications and Networks, DOI: 10.1007/978-1-84882-218-4 4, Springer-Verlag cAst2weszvxvx London Limited 2009.

[17] Manju Bala , Lalit Awasthi "On Proficiency of HEED Protocol with Heterogeneity for Wireless Sensor Networks with BS and Nodes Mobility" International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 2– No.7, May 2012.

[18] Ming Zhang, Yanhong Lu, Chenglong Gong, " Energy-Efficient Routing Protocol based on Clustering and Least Spanning Tree in Wireless Sensor Networks", International Conference on Computer Science and Software Engineering,IEEE, 2008.

[19] Jennifer Yick , Biswanath Mukherjee, Dipak Ghosal" On the Wireless sensor network survey" ELSEVIER, Computer Networks 52 (2008) 2292–2330.

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 3, March 2016

Protection against Phishing in Mobile Phones

Avinash Shende, IT Department, SRM University, Kattankulathur, Chennai, India.
Prof. D. Saveetha, Assistant Professor, IT Department, SRM University, Kattankulathur, Chennai, India.

***Abstract*- Phishing is the attempt to get confidential information such as user-names, credit card details, passwords and pins, often for malicious reasons, by making people believe that they are communicating with legitimate person or identity. In recent years we have seen increase in threat of phishing on mobile phones. In fact, mobile phone phishing is more dangerous than phishing on desktop because of limitations of mobile phones like mobile user habits and small screen. Existing mechanism made for detecting phishing attacks on computers are not able to avoid phishing attacks on mobile devices. We present an anti-phishing mechanism for mobile devices. Our solution verifies if webpages is legitimate or not by comparing the actual identity of webpage with the claimed identity of the webpage. We will use OCR tool to find the identity claimed by the webpage.**

I. INTRODUCTION

Phishing is a criminally fraudulent process of attempting to get sensitive data like usernames, credit card details, passwords and pins, often for criminal reasons, by making people believe that they are communicating with legitimate person or identity [1]. Phishing attacks have seen alarming increase in both volume and sophistication. In response to these threats, researchers have developed various solutions for anti-phishing. There are many anti-phishing schemes present but phishing attacks still continue to happen.

A phishing technique was described in detail in 1987, and the first use of the term “Phishing” was made in 1995. The term is a variation of word fishing, influenced by phreaking, and alludes to “baits” used in hope that the potential victim will “bite” by opening a malicious attachment or opening a malicious link, in which case their financial data and credential will be stolen.

The harm brought by phishing range from denial of access to email, or account or to create huge monetary loss of an individual or the organization. It is estimated that between May 2004 and May 2005, approximately 1.2million computer users in the United States suffered losses caused by phishing, the cost totaling approximately US $929 million. United States businesses loss an estimated US $2 billion per year as their clients become victims. In 2007, phishing attacks escalated to 3.6 million, people lost US$3.2 billion in the 12 months ending in August 2007.In the United Kingdom losses from web banking fraud—mostly from phishing—almost doubled to £23.2m in 2005, from £12.2m in 2004, while 1 in 20 computer users claimed to have lost out to phishing in 2005.According to 3rd Microsoft Computing Safer Index Report released in February 2014, the annual worldwide impact of phishing could be as high as $5 billion. The position embraced by the UK banking body APACS is that "customers must also take sensible precautions ... so that they are not vulnerable to the crimes." Similarly, when the first instant of phishing attacks hit the Irish Republic's banking sector in September 2006, the Bank of Ireland first declined to cover misfortunes endured by its clients. So it becomes customers’ responsibility to keep themselves safe from such kind of Phishing Attacks.

Few years before phishing was done only on desktop sites. After people stated using mobile devices for their online accounts and transactions, the phishers started making phishing pages for mobile phones. When we do internet surfing from a mobile phone, the browser on mobile device will browse for mobile sites which are specially made for mobile devices. Mobile sites are a copy of your website. These sites are very lightweight in nature especially made for small screen devices. These sites have less text content than the desktop pages, and also contain less or no graphics. They are very simple and have different layout than the corresponding desktop pages. In 2012, scientists from Trend Micro discovered 4,000 phishing URLs intended for mobile website pages [2]. Despite the fact that this number is under 1% of all phishing URLs, it highlights that mobile phones have turned out to be new focus of phishing attacks. Phishing attacks in mobiles are easy to perform due to limitation of mobile phones like small screen. Due to small screen of mobile phones it is not possible for user to see complete URL of the webpages. Attackers use this limitation of mobiles for their phishing attacks.

Current phishing page detection mechanism can be divided into two categories: Blacklist based and Heuristics based mechanism [3]. Blacklist based mechanism have already known attacks listed in them, so they can detect phishing websites that are in the black-list but it can't detect zero day phishing attacks which are active for days. If a new phishing site appears, blacklisted method is unable to detect them. Heuristics based mechanism fully depends on key features extracted from HTML source code and URL, and after that different strategies, for example, machine learning is used to decide the legitimacy of the webpage. But we find that some features extracted from HTML source code and URL can be wrong, and phishing websites can easily bypass those heuristics. So we cannot completely depend on either black-listed or heuristics method for detecting phishing attacks.

We propose an anti-phishing mechanism for mobile phones which will detect mobile phishing pages which ask you your confidential information. We believe it is easy to create phishing pages which can bypass heuristics based method which depends on HTML code of the webpage. To avoid that we use OCR tool to get text from the image of login pages. OCR is the electronic or mechanical conversion of image typed, handwritten or printed text into machine-encoded text [4]. The working of OCR is as follows. First the image may have handwritten text or printed text is given as input to OCR tool. Now OCR tool will process on the image and will give us the text present in image in machine-encoded form. OCR method has good performance on mobile devices rather than Computer Systems. We are able to find the claimed identity from extracted text, and actual identity from the URL of the web page. If these two identities are similar, it means that the page is safe. If the identities are different the page will be a phishing page and a warning will be given to the user about it.

II. BACKGROUND OF MOBILE PHISHING

Mobile devices have a very small screen. These devices browse for mobile websites if available, if mobile version of website is not available then it displays normal desktop site. A mobile website is a separate version of a desktop website and it is designed to be used exclusively on smartphone devices. The mobile version of webpage is a limited version of the pages that are displayed on your desktop. When a smartphone user comes to a website, an "auto-detect" will recognize the device they are using and then send that visitor to the mobile version, if the user is using a mobile device to browse. Due to small screen size, most browsers in mobile phones do not display the URL bar when the web page is done loading. While loading process of URL, long URLs are truncated to fit the mobile browser screen. Since the only difference between a Phishing page and a Legitimate page is the URL. It becomes important to see the URL of the page. One method to do so is to scroll through the URL manually. But it takes time and is not very reliable. Figure 1 show how the URL gets truncated.

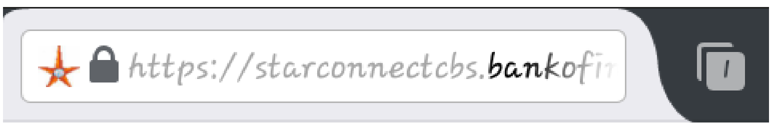


Figure. 1. This figure displays the truncation of URL while a page is loaded in browser.

Now looking at the URL shown in image how can we say that the URL is https://starconnectcbs.bankofindia.com or https://starconnectcbs.bankofindiana.phishing.com. This kind of tricks will fail if complete URL or at least the domain name of the webpage is visible. But unfortunately we are unable to see the complete URL of the Page.

For few websites, their domain name can be easily mimicked by changing or replacing few letters. For example, http://www.srmunív.ac.in. In this URL ‘i’ is replaced by ‘í’ instead of http://www.srmuniv.ac.in. This is very hard to find out while browsing through mobile devices. A user can easily believe that he is browsing on legitimate website. Similarly the small letter ‘i’ can be replaced with capital letter ‘l’. e.g. http://www.srmunlv.ac.in.

Heuristics based mechanism completely depend on features taken from HTML source code and URL, and after that different strategies, for example, machine learning is used to decide the legitimacy of the webpage. But we find that some features extracted from HTML source code and URL can be wrong, and phishing websites can easily bypass these heuristics. Attackers can include text, pictures, and links into HTML code, and at the same time can make “undesirable” things hidden from a webpage by changing its size to zero or putting it behind other picture. So, features like distinct words and there occurrence, brand name, and company logo can easily be misguiding. For example, in figure which displays HTML Source Code.

```
<div class="cl lgoCl">
  <a href="http://kimxydsy.bugs3.com/1.html" _sp='p2054029.m2428.14282'>
    <img class="headerLogo" alt="eBay" src="http://kimxydsy.bugs3.com/Square.jpg"
         style="position:relative; width:0px; height:0px;">
  </a>
</div>

<h1 style="font-size:0%">bugs3 bugs3 bugs3</h1>

<div style="font-size:0%">
  <a href="http://kimxydsy.bugs3.com/1.html">bugs3</a>
</div>
```

Figure. 2. This figure displays the HTML code of a Phishing page which contains “bugs3” as hidden text.

Here the bugs3 is the word which will not be visible to the user on the webpage. Since its font size is zero, it’s not visible on the webpage. But large number of “bug3” will be retrieved. This will make the identity extractor think that this webpage is of “bugs3” and not of “EBay”. So this method will fail since the webpage belongs to “bugs3.com”domain. We showed that the HTML code cannot be dependent on to know the claimed identity of a webpage. So we must depend on what the user see on the webpage that is the screen displayed in the browser.

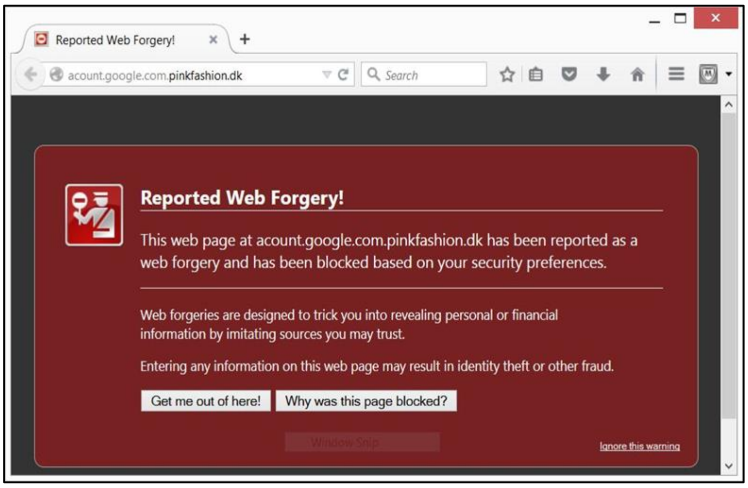


Figure. 3. This image displays the error displayed in a Desktop browser for a phishing page.

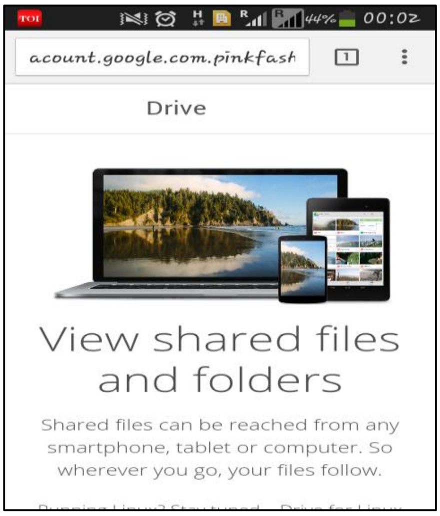


Figure. 4. This image displays the same page in mobile for which Desktop browser displays error.

In Figure 3 and Figure 4 we can see the difference how a browser in Desktop and Mobile device react to a phishing page. The Desktop browser detects the page as phishing page. But the browser in mobile phone do not display any such kind of message. In fact it displays the webpage like any other normal page even if it is a phishing page.

III. OVERVIEW OF OUR ANTI-PHISHING SCHEME

Mostly all organizations use brand name as the second-level domain (SLD) name of their websites. For Example, Bank of America uses the entire brand name as SLD despite its length.

OCR tool converts image into text format. OCR technique gives good performance on mobile devices since screen size of mobile device is small. We use Tesseract OCR. Tesseract is one of the most accurate open source OCR engine. Tesseract is free software, released under the Apache License, Version 2.0. Its development has been sponsored by Google since 2006 [5]. Tesseract version 3 accepts simple one column text as input, multi-columned text, images or equations. Tesseract is suitable for use as a backend. Tesseract does not come with a GUI and is instead run from the command-line interface.

Our solution for Phishing kick starts when the user will start browsing any URL. When the mobile browser tries to load a webpage, we see if its URL is domain name or IP address. Organizations use domain names instead of IP address. Attackers use IP address in URL to hide themselves. Then we obtain the HTML code of the webpage and check if there is any form present in the page. The form is important since attacker need a form which ask user to enter information and then submit. If the form is found, we start identifying the identity of webpage. If form is not available, we stop proceeding further. If form is not present on the page, even though it is a Phishing page it do not cause any harm to user's confidential data like password. If form is detected. We get second level domain name from URL, which tells what website it is. Then we make a whois lookup of the domain from URL [6]. We find which organization has registered the domain. On the other hand, we take an image of a webpage and get the text present in image with help of OCR tool. Then we see if the name of the organization or brand name is present in the text extracted from OCR tool. All the login pages will have the logo of the organization at the top or the copyright statement at the bottom of the page. If the domain from URL is present on the text from OCR tool, we consider the page as legitimate page. If the page is not found to be legitimate, we will warn the user.

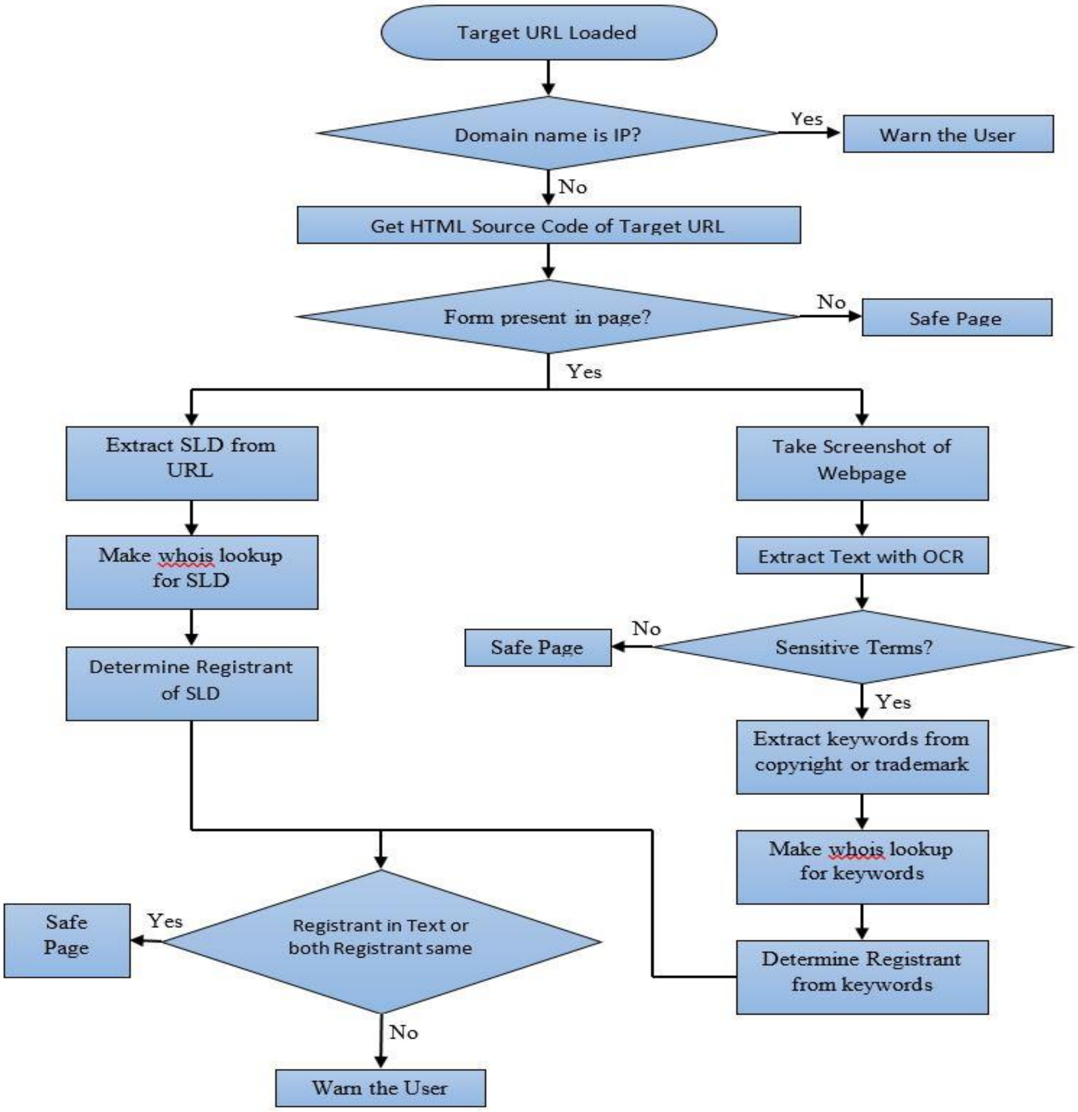


Figure. 5. This image displays the working of our solution.

IV. IMPLEMENTATION

We implement our solution as an Add-On for Nightly Browser in Android [7]. Nightly browser is specially made for developers by Firefox. This browser gets updated on daily basis with new developments from around the world. It is like a testing platform for different modules before bringing them to main Firefox Browser.

One of the important part of this project is to retrieve proper text from the screenshot using OCR Tool. We have used tesseract OCR tool for retrieving text from the screenshot image. Tesseract is one of the best OCR present today. It retrieves text very accurately.

V. CONCLUSION

In this paper, we studied about mobile phishing and its detection mechanisms. We proposed a phishing detection mechanism. We found the weakness of the heuristics based anti-phishing mechanism that depended on the HTML code of the webpage. Our method solves this issue by using OCR tool, which gives text from an image of a login page. This method also works for all domain rather than few selected or whitelisted domains. We implemented this on Samsung Galaxy Grand running the Android 4.2.2 operating system.

REFERENCES

[1] 'Phishing',https://en.wikipedia.org/wiki/Phishing

[2] Trend Micro, "Mobile phishing: A problem on the horizon," 2012.

[3] Longfei Wu, Xiaojiang Du, and Jie Wu: "MobiFish: A Lightweight Anti-Phishing Scheme for Mobile Phone", IEEE, 2014.

[4] 'Optical Caracter Recongnition',https://en.wikipedia.org/wiki/Optical_character_recognition

[5] 'Tesseract OCR', https://github.com/tesseract-ocr/tesseract

[6] 'whois',Who.is/whois/

[7] 'Nightly Browser',https://nightly.mozilla.org/

[8] Hang Xu, Haining Wang, Sushil Jajodia: "Gemini: An Emergency Line of Defense against Phishing Attacks", IEEE Computer Society, 2014, pp 11 – 20.

[9] NourAburaed, HadiOtrok, RabebMizouni, Jamal Bentahar: "Mobile Phishing Attack for Android Platform", 2014, IEEE, pp 18 – 23.

[10] JayshreeHajgude: "Phish Mail Guard: Phishing Mail Detection Technique by using Textual and URL Analysis", IEEE World Congress on Information and Communication Technologies, 2012, pp 297 – 302.

[11] Gang Liu, Bite Qiu, Liu Wenyin: "Automatic Detection of Phishing Target from Phishing Webpage", 2010, IEEE International Conference on Pattern Recognition, pp 4153 – 4156.

[12] Dona Abraham, Nisha S Raj: "Approximate String Matching Algorithm for Phishing Detection", IEEE International Conference on Advances in Computing, Communications and Informatics, 2014, pp 2285 – 2290.

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 3, March 2016

Hybrid Cryptography Technique for Information Systems

Zohair Malki

Faculty of Computer Science and Engineering
Taibah University, Yanbu, Saudi Arabia

Abstract

Information systems based applications are increasing rapidly in many fields including educational, medical, commercial and military areas, which have posed many security and privacy challenges. The key component of any security solution is encryption. Encryption is used to hide the original message or information in a new form that can be retrieved by the authorized users only. Cryptosystems can be divided into two main types: symmetric and asymmetric systems. In this paper we discussed some common systems that belong to both types. Specifically, we will discuss, compare and test the implementation for RSA, RC5, DES, Blowfish and Twofish. Then, a new hybrid system composed of RSA and RC5 is proposed and tested against these two systems when each used alone. The obtained results show that the proposed system achieves better performance.

1. Introduction

Information systems are developing rapidly and their applications exist in almost every aspect of our life. There are many critical fields in which information systems must ensure the security and confidentiality of their systems and ensure that the data and information used in these systems is protected from unauthorized persons. Security is the process used against any malicious or attack; it may detect, prevent the attacks and finally it may recover the attacks effects. The most popular mechanisms of the security are encryption algorithms, digital signature and authentication protocols.

The main security services are: authentication, data confidentiality and data integrity, where the data integrity is ensuring that the message content is not modified through its transferring process; the data confidentiality is the process of ensuring that only the authorized users saw the message; while the authentication service is based on the ensuring that the message sender is a particular user “expected sender”. The cryptography systems are used to apply all of the above services, the data confidentiality, integrity and the user authentication, where it is the process of converting the source message to unreadable form, namely cipher text, in sender side, this process is the encryption, then regenerate the source message by the receiver “decryption process”. The encryption and decryption use specific algorithm and keys; where the keys types and specifications are based on the cryptography algorithm used one secret key, is shared between the two parties, and used in symmetric algorithm while in asymmetric algorithm different keys are used [1]

The public-key cryptosystem is asymmetry cryptography system; these systems use two different keys, one for encryption and one for decryption, namely, private and public keys. Although the symmetric cryptography is faster, the asymmetric systems are more secure and suitable for specific application. Each party using the public-key cryptosystem should have two keys; first one is public key *P* and the second one is the private key *S*. The private key is the secret for the user himself while the public one is known and distributed to anyone needs it by announcement or by publishes it in Public Key Directory (PKD).

The public and private keys are inversed to each other, where if the *P()* is the function related to the public key and S() is the function related to private key. For example, if a sender “Bob” and the receiver “Alice”, when Bob wants to send a message “M” to Alice, Bob will encrypt the message by Alice’s public key before he sends the message, when Alice get the message she will decrypt it using her private key, hence, the public keys is published for everyone. If the message is encrypted by Alice’s private key, it cannot be decrypt by any key other than Alice’s public key, mathematical, $C= S_b(M)$ and $M= P_b(C)$ then $S_b(P_b(M)) = P_b(S_b(M))= M$.

The privacy, authentication and the integrity are main services applied by the public-key cryptography. The privacy service is based on ensuring that is no unauthorized users can know the message content. If Bob sends the message M is encrypted by the public key of Alice, where all public keys are announced by emails or publish in PKD, the cipher text will be unreadable so no one can read it either Alice because she has her secret key. Through the message transferring, if eavesdrop can sniff the message, he cannot read or understand it.

As appeared, that is the privacy by public key cryptography is based mainly on keeping the private key secured. Otherwise, who has the private key can use it to decrypt the message and get it. In public key, we can ensure that is the transferring message will not modify by unauthorized users. If the eavesdropper can sniff the encrypted message, he cannot read or modify it. Even he modifies the message randomly; when the receiver decrypts it the modification will appear so Alice can drop the message and re-request it.

When you send a paper letter, you signed it by your unique signature to make the receiver ensuring about the letter generator. As happened in this case, the digital signature is used to authenticate the two parties to them. The digital signature is different of the encryption-decryption operation. Bob encrypts the message M by Bob’s secret key and concatenates the encrypted message (C) with the source message (M), then sends them to Alice. In Alice side, she will decrypt C by Bob’s public key, then, she will compare M with the decrypted text, if they are equal that means the sender is Bob, else it means a different sender or the packet was altered by eavesdropper or transition errors.

2. Related Work

Depending on the application nature, the needs of any of the above services appears, where some applications focus on the authentication only, other focus on the integrity and privacy and some other can use encryption and digital signature to achieve all of the services. To create public and secret keys we need to select two large (512 bits) and prime numbers p and q where p is not equal to q; then calculate the multiplication of these two prime numbers ($N=p*q$). Now find $Z= (p-1)(q-1)$ and select small odd number (e) where the greatest common divisor between e and Z equal 1 and e is greater than one and less than Z, in this step we found the public key $P=(e, N)$. Then find d while $e*d=1 \bmod Z$ and d is larger than 1 and less than N, hence, the private key is *S(d, N)*. The size of the keys reflects the secure of the system, while if the key size is big then the system will be more secure.

As shown in before section, the public key $P=(e, n)$ is published and used to encrypt the source message while for decryption the secret key $S(d,n)$ will be used, as shown in the below formulas:

$$P(M) = M^e \bmod n = C \quad \text{.... (1)}$$

$$S(C) = C^d \bmod n = M \quad \text{.... (2)}$$

The strength of RSA algorithm is based on the key length and the modulus value (n), on the other hand, these factors are the main reasons of the RSA high runtime complexity. Although the encryption and decryption operations look like the same, the decryption operation with private key takes more time than the encryption operation with public key. The encryption and decryption run time complexities depend on the key used; where private key needs the duplicate number of bits more than the public key.

The RSA system is very secure in case of large size of p and q and on the operation of two random selected integers multiplication, In addition to the security, the RSA cryptosystem does not need to create a new key with every new party and does not need to share it as in the case of symmetric cryptosystems. Finally, the authentication and non-reputation services are provided by RSA system cannot be done by symmetric systems. The main challenge that faces RSA systems is the run speed; the complex computations and large keys generation need enormous time to run. While the larger message means more and more time, a combination of the public key and secret key cryptosystem is proposed to get the advantages of the two systems.

As the improvement on the digital signature, the hash function "*h()*" is used to generate a fingerprint "M'", a fast and easy computational function, where $h(M) = h(M')$. If Bob wish to authenticate himself to Alice, he generate the fingerprint M' by the hash function $h(M)$ to decrease the message size and create un reversible version of the message, then he will encrypt the fingerprint by his secret key $S_B(M')$ and send the encrypted fingerprint and the source message to Alice, send $(S_B(M'), M)$ as the signature.

When Alice get the message she will hash the message $h(M)$ and decrypt the fingerprint by Bob's public key, $P_B(S_B(M'))= M'$. Finally, she will compare the two texts to ensure that the sender is Bob himself, because nobody can create two same fingerprints from two different messages. The main question that still with no answer is how Alice knows that is the published Bob's public key is for Bob really. The certificates are used to distribute the public keys of the users. Assume there is a trusted user T has (P_T , S_T), Alice obtain her public key by get a certificate signed by the S_T. based on the transmission rule, while T trust Alice each one trust T will trust Alice.

The symmetric or the single key cryptosystems are simple and efficient systems use one key on the two parties, the same key is used for encryption and decryption operations. Even though it suffers from many and important problems, these systems are still widely used. Some of the main symmetric algorithms; DES, Blowfish, Twofish and RC5 are discussed next. These algorithms are compared in terms of: time, space, advantages and disadvantages.

The algorithm uses a 64-bit data block and a 56-bit key. DES algorithm is divided into two phases: key expansion and data encryption. The algorithm runs for 16 rounds and a 48-bit sub key is created for each round from the original 56-bit key. The sub-keys are created as follows:

- First, the 56-bit key is used based on the Key permutation table result (PC-1), and then it is divided into 2 equal parts.
- Each part is rotated by 2 bits in each round except for the first, second, ninth and last rounds.
- Compression permutation is used to choose 48 bits from the original 56 bits to form the sub-key.

These sub key are the input of each of the DES round to be used for encryption in the sender side and the similar ones will be used in decryption in the receiver side. Data encryption in DES is divided into three main steps; an initial permutation (IP), 16 rounds of a complex key dependent calculation and a final permutation being the inverse of IP.

The Initial Permutation (IP): the data block is divided into two parts; LH half, and RH half. The block size must be even.

The Data Rounds: takes a 32-bit half (R) and 48-bit sub key and then expands R to 48-bits using expand permutation (E). 8 S-boxes are used (data and sub-key) to get 32-bit result which in turn is permuted using 32-bit perm (P). This process is repeated 16 rounds. In the final round, the left and right halves are not swapped. The result of the last round constitutes the final right half, and the result of the fifteenth round constitutes the final left half.

Final Permutation: The final permutation is the final step, in this stage the final cipher or encrypted data is generated by reorder the input data. The performance of this algorithm is based mainly on the key size, the block size and the arrays search algorithm, where the initial permutation needs constant time while the each round run time is based mainly on the function complexity. Finally, the final permutation needs n times, where n is the size of the block.

As all the symmetric key cryptosystems, DES is fast cryptosystem especially with the small texts. DES's implementation is very easy and simple, mainly because it needs the same encryption and decryption key. DES is insecure because of its small key; 56 bits key can be broken using the brute force attack. It does not solve the authentication and non repudiation problems, so it has been improved to the new and more efficient versions.

The Blowfish is a symmetric cryptography system proposed in [6]. Blowfish is a popular encryption algorithm due to its strength and its free license. This algorithm uses a variable length of key between 32 to 448 bits and 64 bit block size for the encryption operation. Blowfish algorithm is based on simple 16 iterations of the Feistel network, this algorithm is divided into two main parts; the key expansion and data encryption. The Feistel network is a fundamental principle used in many symmetric cryptosystems as DES, Blowfish and Twofish. This principle based on the number of iterations usually the number of iterations is between12 to 16.

Firstly, the block, with N bits size, is divided into two parts L and R eat part has N/2 bits. In each iteration, there is a function f_i where *i* is the current iteration, the input of *f* is the R_{i-1} and the key K_i, the output of the function will be XORed with L_{i-1}, then L and R will be exchanged to generate the input of the next iteration.

Where $R_i = L_{i-1} \oplus f(R_{i-1}, K_i)$ and $L_i = R_{i-1}$

The encoding and the decoding operations are identical in this method so a half of the algorithm complexity is removed. The strength of this principle based on two concepts; the function f_i type and the key k_i value, so the function f should be non linear and the key of the current iteration should be generated dynamically. In this part, many sub keys are generated from the Bluefish key with the size between 32 bits to 448 bits. The sub keys must be pre computed with 32 bits before any encryption or decryption operation using the P-array of 18 entries {P1, P2, ... ,P18}. In addition, 4 by 32-bits S-boxes with 256 entries are used.

The encryption operation contains sixteen Feistel network rounds, where the message is divided into blocks with 64 bits for each, and then each block is divided into two parts with 32 bits for each, L and R. At every iteration, make the previous R is the input to the S-Box with the suitable sub key then XOR-ing the result with L to generate next R and make the next L equal the previous R. Blowfish uses variable size of key between 32 to 448 bits; the runtime of this algorithm is based on the key size and on the used function while the block size is fixed. Overall the Blowfish is faster than DES, where blowfish need approximately 2/3 the DES runtime.

Blowfish is a free license and open source cryptosystem, it is fast and suitable for old systems. While In comparison with other symmetric cryptosystem, Blowfish wastes large size of memory for storing the pre computed sub keys and the S-boxes values. While if we do not pre compute them, the encoding and decoding operations will be slow [5].

This symmetric cryptography algorithm is designed by John Kelsey, Bruce Schneier, David Wagner, Niels Ferguson, Chris Hall and Doug Whiting in 1997. Is the improvement of Blowfish algorithm, also was designed as alternative of DES algorithm with free license and open source for everyone, while Twofish uses 128 bit block and key up to 256 bits. The block is divided here into four parts with 32 bits for each one; one of the two left parts is rotated left eight bits then the output and the other left part will input to the function (S-box), the output of this function is mixes linearly by the MDS matrix. Finally some XORing operations are done on the four parts to get the inputs of the next iteration; this algorithm contains 16 iterations.

Twofish algorithm is rapid in 8-bits and 32-bits CPU and in the hardware and it is secure with strong keys that are up to 256 bits [4]. The Twofish simple design make it is easy to be tested, implemented and analyzed and it has the capabilities to run on different operating systems and different hardware. This algorithm is the free license and open source software, so it is available for anyone with no cost. As any other symmetric key, the secret key exchange is still the main problem, in addition to the multiple keys for single entity [4][12].

RC5 is a block cipher. Designed by Ronald Rivest in 1994, RC stands for "Rivest Cipher", or alternatively, "Ron's Code" it is an improvement of RC2 and RC4. RC5 has a variable block size (32, 64 or 128 bits), key size (0 to 2040 bits) and number of rounds (0 to 255). The block size is 64 bits, a 128-bit key and 12 rounds are used in this algorithm [2].

RC5 uses two parameters: variable block size (w), and variable number of rounds (r). It uses three operations and their inverses as follows:

- Addition and/or subtraction of words.
- Bit-wise exclusive-or (XOR).
- Rotation.

First, the password key (K) is expanded to large size using the expansion table (*S*). Then, in the encryption process, a plain text is transformed to ciphered text. RC5 is a simple and easy to implement algorithm, the runtime depends on the key size, block size and the number of rounds.

3. Proposed Approach

With large message the high encryption and decryption time problem appears, so the hybrid cryptosystem are used. In this system the secret key will be used in addition to the public and private keys. If Bob, has a public and secret keys (P_B, S_B), wants to send a message M to Alice, he will use a symmetric key (K) to encrypt the message, that means fast encryption and decryption, and then he encrypts the symmetric key with Alice's public key (P_A). In other side, when Alice receive the message she will decrypt the encrypted symmetric key by her secret key (S_A) to get the symmetric key to use it to decrypt the source message (Figure 1).

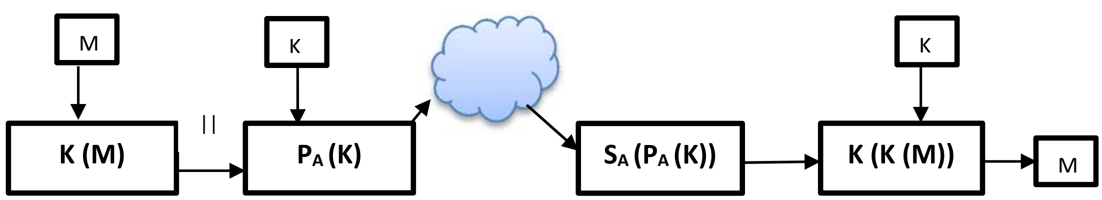


Figure 1: Hybrid Cryptosystem

In [8], a cryptosystem is proposed for key distribution, the DES cryptosystem is used for encryption and decryption operations with 168-bit key, while the public key cryptography system is used for DES key distribution. While in [9], the AES is used to encrypt the secret key that used in the data encrypted by ECC. Then the cipher text and cipher key are sent to the receiver, the receiver decrypts the key using his private key then use the decryption output to decrypt the message. Although the RSA cryptosystem is very secure, has no key exchanging problem and is providing the authentication and non reputation, it is so slow to use with large message. The best solution to get all above advantages of RSA adding to the fastness of the symmetric key cryptosystems, especially the RC5, is to use the hybrid systems.

In these types of systems, the advantages of RSA and the symmetric key cryptosystems are combined. Where Bob, has a public and secret keys (P_B, S_B), wants to send a message M to Alice, he will use a symmetric key (K) to encrypt the message, that means fast encryption and decryption, and then he encrypts the symmetric key with Alice's public key (P_A). In other side, when Alice receive the message she will decrypt the encrypted symmetric key by her secret key (S_A) to get the symmetric key to use it to decrypt the source message (see Figure 2).

Some hybrid systems are proposed and used previously, in [7] the hybrid cryptosystem is proposed for key distribution, the DES cryptosystem is used for encryption and decryption operations with 168-bit key, while the public key cryptography system is used for DES key distribution. While in [8], the AES is used to encrypt the secret key that used in the data encrypted by ECC. Then the cipher text and cipher key are sent to the receiver, the receiver decrypts the key by his private key then use the decryption output to decrypt the message.

As discussed before, asymmetric key cryptosystems as RSA are more secure than the symmetric key cryptosystems, but it has main problem with the run time [3]. While the symmetric key cryptosystems are fast but they have a problem with the key exchanging especially in first communication. After we implemented the main symmetric cryptosystems, we found that RC5 is the best one. So our future work is the RC5-RSA cryptosystem, where the hybrid system consists the RC5 to encrypt the message itself using hashed secret key, hashing the key to reduce its size before encrypt the hashed key by the receiver public key.

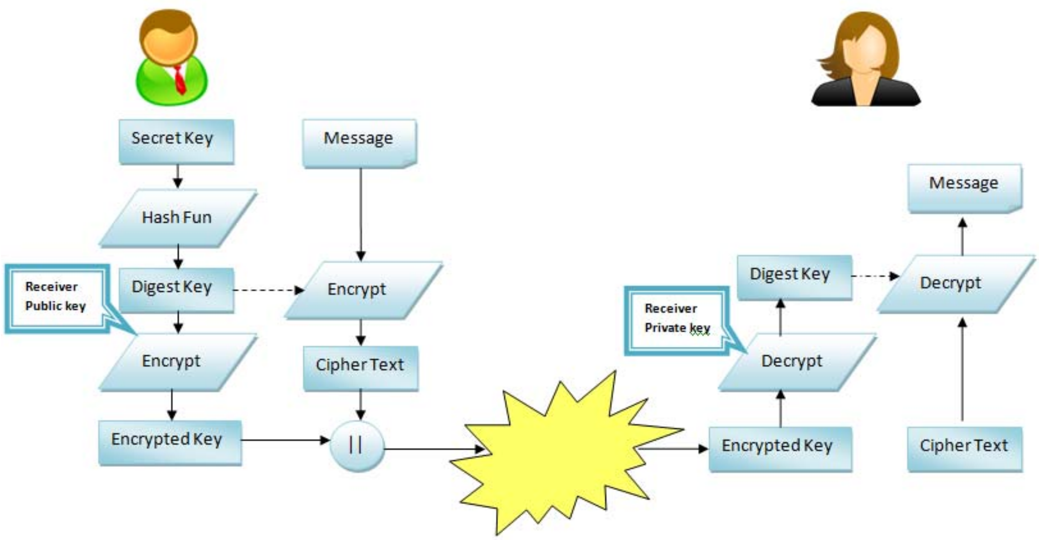


Figure 2: Proposed algorithm

4. Simulation and Results

In the first part of our simulation experiments we have compared the current approaches: RSA, Twofish, Blowfish, DES and RC5. The simulation code have been written in C++ under a PC running Windows 7 OS with Intel Core 2 Processor (2 GHz) and Memory of 2 GB.

These experiments have been repeated for 8 different file sizes starting from 6000 bytes and doubled each time. The comparison criterion is the running time of each experiment. Table 1 show the obtained results which also illustrated in Figure 3.

Table 2: Simulation results for all algorithms

file size (bytes)	RSA	DES	Bluefish	Twofish	RC5

6000	0.993519	1.051524	0.343576	0.422571	0.201413
12000	3.951245	1.537545	0.701244	0.791625	0.251711
24000	6.305834	2.772531	1.097448	1.274363	0.342455
48000	14.49432	6.158043	2.37076	2.914358	0.359375
96000	26.5751	11.69644	4.490399	5.321609	0.408743
192000	47.79881	21.91718	7.55493	9.576195	0.805201
384000	102.6349	48.47278	17.85167	20.53655	1.232432
768000	175.9754	77.019	31.304	35.20102	1.831377

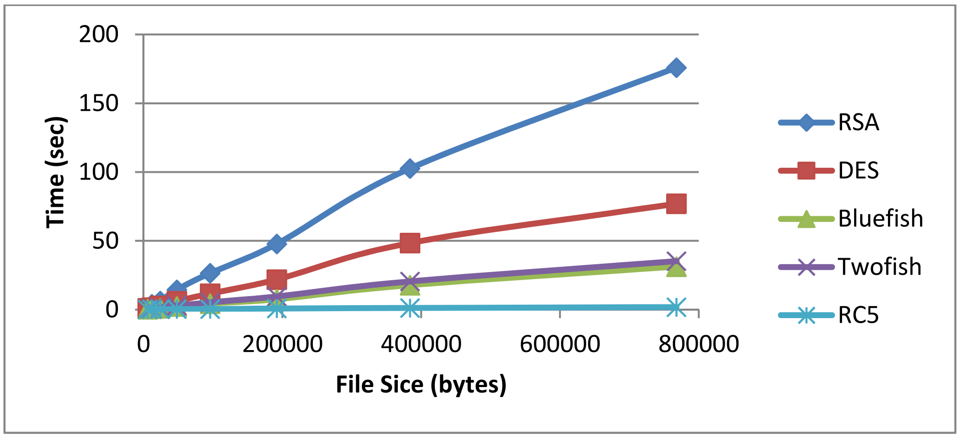


Figure 3: All algorithms running times

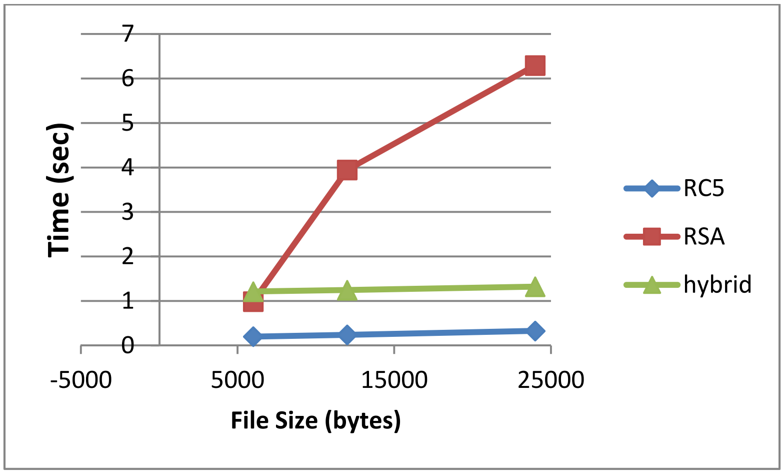
As shown above from the implementation result the RC5 is the best encryption algorithm since it has the best running time for all the input file sizes compared to all other encryption algorithms. In contrast the RSA gives the worst running time result although for small input file size, but still RSA is asymmetric key algorithm so it is secure and it has no problem with keys exchanging. RSA must use with the small block size to get efficient works.

This table shows the result of the implementation of Hybrid algorithm compared to RC5 and RSA. The system was run with deferent ten input files size, each time we duplicate the file size. As we can see in Table 2 the result table and graphs, the running time of hybrid algorithm is very close to the RC5 running time and much better than RSA.

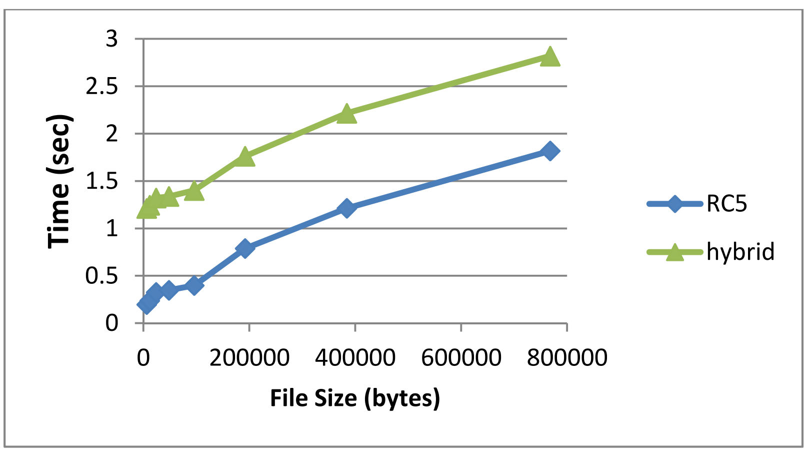
Table 2: Simulation results for RC5, RSA and the Hybrid algorithms

file size (bytes)	RC5	RSA	hybrid
6000	0.198	0.988	1.211
12000	0.238	3.947	1.245
24000	0.326	6.294	1.322

48000	0.348	14.482	1.338
96000	0.398	26.574	1.401
192000	0.788	47.791	1.762
384000	1.213	102.632	2.216
768000	1.818	175.961	2.819



(a) comparing RC5, RSA and the Hybrid algorithms



(b) comparing RC5 and the Hybrid algorithms

Figure 4: RSA, RC5 and Hybrid algorithms running times

We can see in Figure 4(a) the run times for RSA, RC5 and RC5-RSA systems. Since RSA time grows exponentially, the figure has been redrawn for only RC5 and the Hybrid algorithm and for larger file sizes as shown in Figure 4(b).

5. Conclusion

RSA system is public key cryptography system and is used to improve number of security service like the integrity, confidentiality and the privacy. The integrity and confidentiality are done by the public and private keys encryption and decryption while the privacy is applied by the digital signature. RSA cryptosystems strengths are based on the public and private keys long and based on the hard mathematical operations, all of that make these systems are secure and unbreakable. The same factors of the strength make the encryption and decryption operations are slow for the long text, so the hybrid system is used widely. The hybrid systems consist of the secret key cryptography system to encrypt the long text and the public key cryptosystem for encrypting the secret key to get secure transferring for the key. The DES, Blowfish, Twofish and RC5 are symmetric key cryptosystems that use single key for encryption and decryption. The symmetric key cryptosystems are faster than asymmetric cryptosystems but in general they have a key exchanging problem.

The hybrid system is proposed to employ the benefits of symmetric and asymmetric systems together. We implemented the four symmetric algorithms and we found that the RC5 is the fastest one, while RSA is popular, unbroken and secure system.

References

[1] Kessler, Gary C. "An overview of cryptography." (2003).

[2] Singh, Ashmi, Puran Gour, and Braj Bihari Soni. "Analysis of 64-bit RC5 Encryption Algorithm for Pipelined Architecture." International Journal of Computer Applications 96, no. 20 (2014).

[3] Abdalla, Michel, Fabrice Benhamouda, and David Pointcheval. "Public-key encryption indistinguishable under plaintext-checkable attacks." In Public-Key Cryptography--PKC 2015, pp. 332-352. Springer Berlin Heidelberg, 2015.

[4] Schneier, Bruce, John Kelsey, Doug Whiting, David Wagner, Chris Hall, and Niels Ferguson. The Twofish encryption algorithm: a 128-bit block cipher. John Wiley & Sons, Inc., 1999.

[5] Al-Neaimi, Afaf M. Ali, and Rehab F. Hassan. "New Approach for Modifying Blowfish Algorithm by Using Multiple Keys." IJCSNS, March 3 (2011).

[6] Schneier, Bruce. "Description of a new variable-length key, 64-bit block cipher (Blowfish)." In Fast Software Encryption, pp. 191-204. Springer Berlin Heidelberg, 1994.

[7] Kaliski, B. S., and Yiqun Lisa Yin. On the security of the RC5 encryption algorithm. RSA Laboratories Technical Report TR-602. 1998.

[8] Parnerkar, Amit, Dennis Guster, and Jayantha Herath. "Secret key distribution protocol using public key cryptography." Journal of Computing Sciences in Colleges 19, no. 1 (2003): 182-193.

[9] Balitanas, Maricel O. "Wi Fi protected access-pre-shared key hybrid algorithm." International Journal of Advanced Science and Technology 12 (2009): 35-44.

[10] Rivest, Ronald L. "The RC5 encryption algorithm." In Fast Software Encryption, pp. 86-96. Springer Berlin Heidelberg, 1995.

[11] Fujita, Takahiro, Kiminao Kogiso, Kenji Sawada, and Seiichi Shin. "Security enhancements of networked control systems using RSA public-key cryptosystem." In Control Conference (ASCC), 2015 10th Asian, pp. 1-6. IEEE, 2015.

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 3, March 2016

AN EFFICIENT NETWORK TRAFFIC FILTERING THAT RECOGNIZE ANOMALIES WITH MINIMUM ERROR RECEIVED

Mohammed N. Abdul Wahid and Azizol Bin Abdullah
Department of Communication Technology and Networks,
Faculty of Computer Science and Information
Technology, University Putra Malaysia, Malaysia

Abstract

The main method is related to processing and filtering data packets on a network system and, more specifically, analyzing data packets transmitted on a regular speed communications links for errors and attackers' detection and signal integrity analysis. The idea of this research is to use flexible packet filtering which is a combination of both the static and dynamic packet filtering with the margin of support vector machine. Many experiments have been conducted in order to investigate the performance of the proposed schemes and comparing them with recent software's that is most relatively to our proposed method that measuring the bandwidth, time, speed and errors. These experiments are performed and examined under different network environments and circumstances. The comparison has been done and results proved that our method gives less error received from the total analyzed packets.

Keywords: *Anomaly Detection, Data Mining, Data Processing, Flexible Packet Filtering, Misuse Detection, Network Traffic Analyzer, Packet sniffer, Support Vector Machine, Traffic Signature Matching, User Profile Filter.*

1 *Introduction*

The advantage of network traffic analysis is that it can monitor the network traffic of local area network and helping discover network problems and alert users when attacker's behavior appears. So, it can protect the network and reads the traffic routs from the source up to destination which connected to that network also from being penetrated by unauthorized user, so it has the advantages of capturing both the normal and the abnormal traffics. A packet sniffer, sometimes referred to as a network monitor or network analyzer, can be used legitimately by a network or system administrator to monitor and troubleshoot network traffic. Using the information captured by the packet sniffer an administrator can identify erroneous packets and use the data to pinpoint bottlenecks and help maintain efficient network data transmission. In its simple form a packet sniffer simply captures all of the packets of data that pass through a given network interface. Typically, the packet sniffer would only capture packets that were intended for the machine in question. However, if placed into promiscuous mode, the packet sniffer is also capable of capturing all packets traversing the network regardless of destination. A packet sniffer can only capture packet information within a given subnet. So, it's not possible for a malicious attacker to place a packet sniffer on their home ISP network and capture network traffic from inside your corporate network (although there are ways that exist to more or less "hijack" services running on your internal network to effectively perform packet sniffing from a remote location). In order to do so, the packet sniffer needs to be running on a computer that is inside the corporate network as well. However, if one machine on the internal network becomes compromised through a Trojan or other security breach, the intruder could run a packet sniffer from that machine and use the captured username and password information to compromise other machines on the network.

This paper explaining the essence of network traffic analysis and its ability to capture and monitor all the traffics (incoming and outgoing), also the ability of detecting suspicious activities that targeting the network such as intruder pushing unwanted programs to degrade the performance. The most widely deployed methods for detecting cyber terrorist attacks and protecting against cyber terrorism employ signature based detection techniques. Such methods can only detect previously known attacks that have a corresponding signature, since the signature database has to be manually revised for each new type of attack that is discovered. These limitations have led to an increasing interest in intrusion detection techniques based on data mining [2, 7]. Data mining based intrusion detection techniques generally fall into one of two categories; misuse detection and anomaly detection. In misuse detection, each instance in a data set is labeled as 'normal' or 'intrusive' and a learning algorithm is trained over the labeled data. These techniques are able to automatically retrain intrusion detection models on different input data that include new types of attacks, as long as they have been labeled appropriately. A key advantage of misuse detection techniques is their high degree of accuracy in detecting known attacks and their variations. Their obvious drawback is the inability to detect attacks whose instances have not yet been observed. Anomaly detection approaches, on the other hand, build models of normal data and detect deviations from the normal model in observed data. Anomaly detection applied to intrusion detection and computer security has been an active area of research since it was originally proposed by Denning. Anomaly detection algorithms have the advantage that they can detect new types of intrusions as deviations from normal usage [1, 2]. In this problem, given a set of normal data to train form, and given a new piece of test data, the goal of the intrusion detection algorithm is to determine whether the test data belong to "normal" or to an anomalous behavior. However, anomaly detection schemes suffer from a high rate of false alarms. This occurs primarily because previously unseen system behaviors are also recognized as anomalies, and hence flagged as potential intrusions. It is very difficult to set any predefined rule for identifying correctly attack traffics since there is no major difference between normal and attack traffic. So, the problem is the fundamental difficulties in achieving an accurate declaration of an intrusion to solve the problem of the high rate of false positive alarm. Also users may slowly change their behavior with the system and time evolution (e.g. the traffic in a network may present changes and variations), and therefore, any associated algorithm should be capable of dynamically adapting to these changes and evolutions.

This paper emphasizes on the design and development an enhanced strategy that can be used to improve the accuracy of the prediction of the network traffic normality, in practical, this paper focus on anomaly detection based on flow monitoring and as a result of the overall anomaly detection methodology, especially in cases where high burstiness is present. First of all, the author proposing a mechanism that provides effective traffic separation and filtering based on 'frequency domain' to analyze the captured network traffics. This approach is based on the observation or the dataset that the various network traffic components, are better identified, represented and isolated in the frequency domain. Specifically, when separating the traffics into two main components: the baseline component and the short term component. The mechanism of packet filtering that analyzes and detects suspicious activity simultaneously for local area network. The new structure design of packet filtering that allows users to capture and detect any intruders that may interrupt or compromise our network, the modifying the basic concept of network traffic analysis has been made to come out with new generation and apply new algorithm into the structure of packet filter, such as traffic signature matching (TSM) and traffic source separation (TSS). Each one of these functions has a specific mission; this mission has to be achieved during the packet filtering. Network analyzers have used three types of network filter:

1-Traffic Filtering: Traffic filtering is a method used to enhance network security by filtering network traffic based on many types of criteria.

2-Packet Filtering: Packet filtering is a method of enhancing network security by examining network packets as they pass through routers or a firewall and determining whether to pass them on or what else to do with them. Packets may be filtered based on their protocol, sending or receiving port, sending or receiving IP address, or the value of some status bits in the packet. There are two types of packet filtering. One is static and the other is dynamic. Dynamic is more flexible and secure as stated below.

Static Packet Filtering: This filter does not track the state of network packets and does not know whether a packet is the first, a middle packet or the last packet. It does not know if the traffic is associated with a response to a request or is the start of a request.

Dynamic Packet Filtering: This filter tracks the state of connections to tell if someone is trying to fool the firewall or router. Dynamic filtering is especially important when UDP traffic is allowed to be passed. It can tell if traffic is associated with a response or request. This type of filtering is much more secure than static packet filtering.

3-Flexible Filters: network analyzer can filter out all types of packets that are coming from different types of network. So flexible filter can filter traffics by:

- Flexible filter: Packet Filter, Email Filter, Web Access Filter
- By MAC address or IP address
- By port numbers
- By protocols
- By packet size, packet value or packet pattern
- Advanced Boolean rules for complex filter formulas (Enterprise edition)
- Supports multi filters simultaneously
- Tracks filter history
- Shares filter settings between projects

So by using this type of filter, administrators can get much more details about the packets and also there is no packet will be unfiltered or the filtering status is incomplete because the flexible filter has the ability to filter out all types of network traffics that are traversing over the local area network. The most important point of the flexible packet filtering is that it takes the advantages of using the attributes of both the dynamic and static packet filtering and it has the ability of tracing and matching requests and replies, the dynamic filter will identify the reply that does not match a request, when the attacker trying to penetrate the filter by making packet looks like reply packet, this can be done by indicating reply in the header of the packet. When the request is recorded the dynamic filter open small inbound hole, so, only the expected data reply is let back through. Once the reply is received the small inbound hole will close.

The proposed robust method that detects network anomalous traffic data based on flow monitoring. This method works based on monitoring the four predefined metrics that capture the flow statistics of the network [10, 23]. In order to prove the power of the new method, an application that detects network anomalies has been build to support the proposed method. And the result of the experiments proves that by using the four simple metrics from the flow data, the system do not only effectively detect but can also identify the network traffic anomalies. Internet traffic measurement is essential for monitoring trends, network planning and anomaly traffic detection. In general, simple packet- or byte-counting methods with SNMP have been widely used for easy and useful network administration. In addition, the passive traffic measurement approach that collects and analyzes packets at routers or dedicated machines is also popular. However, traffic measurement will be more difficult in the next-generation Internet with the features of high-speed links or new protocols such as IPv6 or MIPv6.

The main contributions of this paper can be explained as the following:

1. An efficient method that enhance the detection of network anomalies by combining special attributes of the static and dynamic packet filtering into flexible packet filtering of the network traffics analysis.

2. An improved algorithm of SVM is proposed to fit the requirements of the network traffics analysis software to be working upon the flexible packet filtering to provide efficient anomaly detection and traffic classification with minimizing the rate of false alarm.
3. An improved method will be applied on top of the network analyzer to make it able for dynamically adapting to the change of user behavior as well as new testing method that could expose any analyzer software to be tested against attacks or any other threats that users might face while surfing the internet.

The proposed method is actually depending on the network traffic analyzer to capture and analyze the network traffics and this research have proposed a special technique that detects anomalies while monitoring network traffics, this technique is called the Flexible Packet Filtering of the Support Vector Machine. So, this research have merged the analyzed results for both of flexible packet filtering and SVM algorithm that used to get better classification of the captured network traffics and to detect anomalies. The idea is to use flexible packet filtering to filter out the captured network traffic and will use the User Profile Filter UPF that will be based on Support Vector Machine (SVM) to detect an attacks that caused by known users and to trace the source of the suspected packet using IP trace back that will be based on traffic source separation TSS of the network traffic monitoring for log-based trace-back. The most important issues of the flexible packet filtering is that it takes the advantages of using the attributes of both the dynamic and the static packet filtering and it has the ability of tracing and matching requests and replies, the dynamic filter will identify the reply that does not match a request, when the attacker trying to penetrate the filter by making packet looks like reply packet, this can be done by indicating reply in the header of the packet. When the request is recorded the dynamic filter open small inbound hole, so, only the expected data reply is let back through. Once the reply is received the small inbound hole will close.

2 Related Works

In [33] a strategy that effectively combined strategies of data mining and expert system was used to design an Intrusion Detection System (IDS). This technique has appeared to be promising but there are some problems in structural and the system performance. In addition, combining multiple techniques in designing the IDS is a recent event and it needs further improvement. The signal analysis approach in Ref. [28] takes an approach that is distantly related to the strategy, used in this paper, of using periodic functions to approximate the reconstituted time series. It applies wavelet analysis to an unstructured time series; this approach has become popular in the analysis of self-similar time series. However, the analysis is an offline technique and did not yield clear advantages over cheaper methods. The need for a policy for resolving subjective ambiguities in computer systems has been explored in a variety of access-security related contexts [25, 26], but this is not a concept that has been discussed for pattern recognition. For intrusion and anomaly systems, policy usually amounts to defining lists of regular expressions to match symbolic traffic payloads. Although not all symbolic languages are regular, any finite symbolic language is regular [13] and all sequences are finite in practice. The computational simplicity of using regular expressions makes this approach the overwhelming approach of choice. Policy is normally only applied to Intrusion Detection Systems and firewalls, rather than anomaly detection systems; see for example Ref. [11]. Approaches that attempt to characterize and utilize the shape of statistical distributions, other than implicitly with a Gaussian model, are unknown to the present author.

Traffic analysis and monitoring (TAaM) are important for internet management and have been widely studied in recent years. By analyzing the IP address attributes, many interesting findings are put forward for abnormal behavior detection. In [16], traffic packets are projected to four matrices according to different bytes of the IP address, and then an abnormality detection method for large scale network is proposed. The structure of addresses contained in IPv4 traffic with different length of prefixes is analyzed and many interesting findings are proposed for traffic measurement and monitoring in Kohler et al, [17]**.** Besides analyzing the characteristic of the IP address, many researchers try to discover the statistical character of users' behaviors and perform abnormal behavior detection [24, 34]. The protocol, client, server port, total data transferred are used to describe the users' communication patterns and to cluster them into different community of interests (COI). Through analyzing the characteristics of the COIs, many abnormal behavior detection methods are designed. The abnormal behaviors can be detected by analyzing the protocols, packets size and flow size [14]. In implementing this idea, it is usually required to check every packet to get the detailed address information. In actual application, this may affect the efficiency of the real-time traffic monitoring. To avoid this and improve the efficiency, some researchers analyze the statistics of the traffic packets (total number of bytes, total number of packets, etc.) and successfully propose many schemes to discover the anomalies only when traffic pattern changes due to attacks (such as DDOS) [12, 23]. The basic idea of those methods is to establish a statistical profile of normal behaviors and then check the current traffic patterns to detect any abnormal behavior. This idea is very useful for detect large scale abnormal behaviors. However, with the increasing number of the internet users and bandwidth usage, detecting abnormal behaviors by just analyzing the characteristic of the total number of traffic packet (or byte) would not be effective since many abnormal behaviors would not cause significant changes in the traffic volumes. To overcome this difficulty, the NetFlow model is proposed by CISCO [4, 5], where a traffic flow is defined as a group of packets with the same source and destination IP addresses ports, etc. NetFlow model is widely used in traffic monitoring systems and many abnormal behavior detection methods are designed based on the signal processing techniques [35]. One of them is the wavelets methods, which are mathematical functions that cut up traffic data into different frequency components, and then the anomalies can be detected by examining the mid- and high-frequency components. Another type of methods is the time-series forecasting methods, e.g. the exponentially weighted moving average (EWMA) method [36]. The EWMA control charts are proved to be a good estimation even under ill conditions and can be used to detect the changes and abnormal behaviors. By setting the lower and the upper limits, the abnormal behaviors can be detected if the monitored feature falls outside the ranges. In fact the idea of these methods is to detect deviations from an expected norm. Generally, the distributions of flow features are quite stable in a monitoring time window. The abnormal network behaviors would cause changes of the distributions and be measured by entropy. The entropy based methods are developed to measure feature distributions and detect anomalies that cannot be identified by the volume based analysis alone [19, 20, 22 and 27]. Many abnormal behaviors with specific traffic flow patterns can be detected by setting a threshold on the number of specific type flows, e.g., the number of ICMP packets for detecting worms [15, 31 and 38]. Since the NetFlow model is a one-way flow model, it may not capture the interactive traffic characteristics. To overcome this shortcoming, a bidirectional flow model is proposed and widely used for traffic classification [18, 21]. One of the major difficulties with above traffic flow models is that the number of flow records could be huge and serious computational and storage difficulties may be encountered. There are mainly two kinds of methods in the literature to extract or aggregate flow information. Sampling can greatly reduce the computational complexity by storing and processing only a very small subset of network packets. Many sampling methods are proposed for high speed traffic monitoring, such as random packet sampling, smart sampling and sample-and-hold [10, 37]. However, some important flow fingerprints would be missed in sampling especially when a large number of flows mixed with only a small number of flows generated by

DoS and D-DoS attacks [3]. The sketch method is another method for analyzing massive data streams. It is based on a probabilistic dimension reduction technique that ‘‘sketches’’ a huge number of per-flow states into a probabilistic summary for massive data streams. The sketch method has successfully applied been for detecting heavy hitters or changes and estimating flow size distribution, which are critical for network traffic monitoring, accounting and anomaly detection [7, 32]. However, this method may encounter the same issue of missing desired flow signatures as sampling methods do.

Intrusion Detection System (IDS) mainly uses two types of techniques, signature based intrusion detection system and anomaly based [9]. Signature based IDS uses predetermined and pre-configured rules or signatures to identify traffic as attack traffic or legitimate traffic and second is anomaly based intrusion system, it refers to the problem of finding patterns in the traffic data that do not behave as expected and alarms an attack if there is abnormal behavior in the traffic pattern. Problem with signature based intrusion detection system is that it can only detect the attacks of which it has rules. On the contrary anomaly based intrusion detection system can detect a new attack with the assumption that at the time of attack, network behavior changes [29]. Anomaly based intrusion detection system uses entropy values of different network features and different data mining techniques [19, 20]. The entropy of different network feature attributes is observed under normal and abnormal network conditions [23]. A hybrid method is also proposed for anomaly detection [30], combining both techniques that are entropy and support vector machine (EaSVM). Firstly, Normalized entropy values of different network features are calculated. Then SVM model is trained in order to classify the normal traffic vs. attack traffic. To understand and evaluate the anomaly traffic detection techniques, second week of traffic data provided by MIT Lincoln Laboratory are used (DARPA, 1999) in [6].

3 *Materials and Proposed Method*

This paper focuses on improving the detection of novel attacks by using new concept of network packet filtering for network traffic analysis system. To achieve the proposed idea, there should be steps to follow to satisfy the research objectives. These steps have some problems and limitations that have been detected in the previous scheme and we have come out with an improved scheme design, analytical study and experimental simulation that overcome these problems and limitations, and to evaluate the experiment, results and performance evaluations must be compared with other systems that are most relevant to this proposed method.

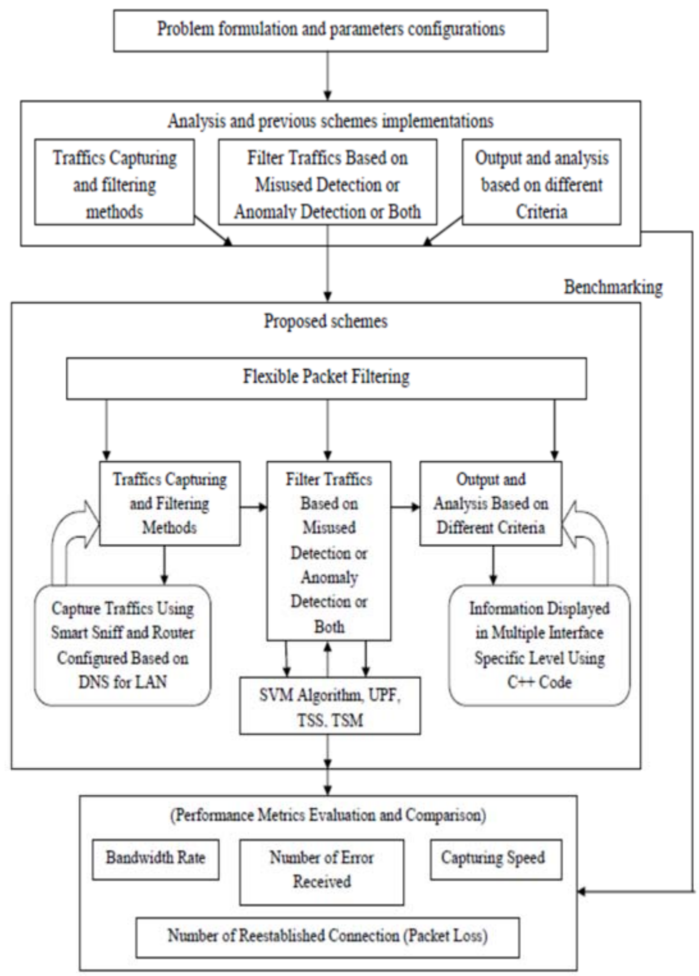


Figure 1; system design as research frame work

This study starts with intensive review of the existing literature for more than one focus point in this area to have full image about the problems that users might face during the implementation, and these literature came from studying many research articles and resources to collect the required information about the pervious and the current problems and to analyze the proposed solutions.

So, this paper focuses on three main issues:

- Traffics filtering methods for anomaly detection in LAN using network traffics analysis.
- Reducing the false alarm rate by using the Support Vector Machine algorithm over the User Profile Filter to detect abnormal behavior from known or unknown users.
- The research also focus on displaying the entire traffics information in a high level interface with an improved functions that allow user easily monitor and troubleshot the network traffics problems, and the testing has been done using network traffics generator.

The idea of this technique is to use the maximize margin of the support vector machine along with the user profile filter that would work perfectly for indentifying and alerting against abnormal activities while monitoring the network traffics. The use of SVM in our proposed method is the same with the EaSVM [2, 30] that most anomaly detection algorithms require a set of purely normal data to train the model, and they implicitly assume that anomalies can be treated as patterns not observed before. The second idea that completes the dimensions of this research is to use flexible packet filtering (FPF) which is a combination of both the static and dynamic packet filtering in the network traffic analysis to filter out the captured network traffic. After that all the captured traffics will be isolated based on their source using traffic source separation 'TSS' strategy that works based on local DNS and during the separation operation the traffic signature will be examined with the stored signatures of the system database using Traffic Signature Matching. After that will create a User Profile Filter (UPF) that will be based on SVM that have the record of the normal users' activities on same work group or DNS for the local area network that must be analyzed to classify the captured network traffics into normal or attack.

As we know that the DNS is a Domain Naming System and the Domain is a group of users and computers managed by the same security database. This paper focusing on these five performance metrics for network evaluation;

- Flow monitoring
- Availability
- Loss & Error
- Delay
- Bandwidth

Flow monitoring is one of the most important metrics that should be considered in network analyzer to capture the flow of traffics from the network then other metrics will show more information and results of the flow monitoring. Flow monitoring usually can adapt both Router based Network Analyzer and non-Router based Network Analyzer. Availability metrics assess how robust the network is, i.e. the percentage of time the network is running without any problem impacting the availability of services. It can also be referred to specific network elements (e.g. a link or a node), and in that case it will measure the percentage of time they are running without failure. Loss and error metrics are indicative of the network congestion conditions and/or transmission errors and/or equipment malfunctioning. They usually measure the fraction of packets lost in a network due to buffer overflows or other reasons, or the fraction of error bits or packets.

Delay metrics also assess the network congestion conditions or effect of routing changes. They measure the delay (One Way Delay-OWD and Round Trip Time-RTT) and Delay Variation (IPDV, or "jitter") of the packets transferred by a network. Finally, bandwidth metrics assess the amount of data that a user can transfer through the network in a time unit, both dependent and independent from the existing network. Bandwidth requirements vary from one network to another. Determining how many bits per second travel across the network and the amount of bandwidth each application uses are vital to build and maintaining a fast, functional network.

4 *Flexible Packet Filtering (FPF)*

The main goal of the Flexible Packet Filtering is to enhance the stage performance and the ability of network analysis system for detecting anomalies and alert users to their presence. These enhancements can be exhibited by means of detecting new attacks and decreasing the false alarm rate that indicating anomalies while monitoring the network traffics. These enhancements can be done by using new strategies upon the following steps:

1. **Capture Traffics**: Traffics are captured based on selected website; it might be multiple websites with multiple protocols.
2. **Filter Traffics**. Traffics are filtered based on the following techniques:

a- *Flexible **P**acket **F**iltering* (FPF)
To identify types of attack using our proposed FPF, we focused on traffic four metrics that each one of them indicating a different type of attack.

- Total Byte
- Total Packet
- D-Socket
- D-Port

b- *Traffic **S**ignature **M**atching* (TSM)
c- *User **P**rofile **F**ilter* (UPF)
d- **Classification Based on Parameters:** Traffics are classified based on their source using **T**raffic **S**ource **S**eparation (TSS) and some other parameters.
e- **Results and Analysis:** Using specific parameters for deeply analyzing the network traffics to give accurate detailed information about the nature of the traffic as a final result.
The abnormal behaviors are usually defined as incidents affecting normal Internet operation such as those aiming to compromise or disable hosts or networks. Table 1 listing a set of anomalies commonly encountered in backbone network traffics.

Anomaly type	Anomaly traffic characteristics
Port scan	Probes to many destination ports
Network scan	Probes to many destination addresses
Worms	Scanning by worms for vulnerable host
Alpha flows	Unusually large volume from point to point
DOS	Unusually large volume, e.g. SYN-Flood
Flash crowd	Burst of traffic to single destination
Outage events	Traffic shifts due to equipment failures
Content distribution	Many traffic volumes from single source to many destinations

Table1, Qualitative with major effects on traffic patterns by various anomalies.

5 *Proposed Filtering Technique (FPF) with SVM Algorithm*

Support vector machine is one of the useful classification techniques [58, 61]. SVM model is learned using different network features for classifying the normal traffic and attack traffic. After training of SVM model using network features, it will be able to predict whether the traffic falls into the one category that is attack traffic or the normal traffic .The support vector machine algorithm creates normal profile and helps to flags data whether normal or anomaly, and returned data will be added to the oldest pattern and constructs new updated normal profile that contain more information about normal user behavior.
Below is the theory formula of the proposed SVM; we have L training points, where each input (xi) has D attributes (i.e. is of dimensionality D) and is in one of two classes' y_i = -1 or +1, i.e our training data is of the form:

$$\{Xi, yi\} \text{ where } i = 1...L, yi \in \{-1,1\}, X \in \Re^D$$

Here we assume the data is linearly separable, meaning that we can draw a line on a graph of X_1 vs. X_2 separating the two classes when D = 2 and a hyper-plane on graphs of X_1; X_2 : : : X_D for when D > 2.
This hyper-plane can be described by w. x + b = 0 where: w is normal to the hyper-plane, b over ||w|| is the perpendicular distance from the hyper-plane

to the origin. So, the final form of the SVM that we used is: **f(x) = w x + b.**
Using SVM Algorithm to Create a User Profile and to Detect Anomalies**;** In SVM, if we define the distance from the separating hyper-plan to the nearest expression vector as the margin of the hyper-plane, then the SVM selects the maximum margin separating hyper-plane. Selecting this particular hyper-plane maximizes the SVM s ability to predict the correct classification of unseen pattern. And the graph shows how SVM works.

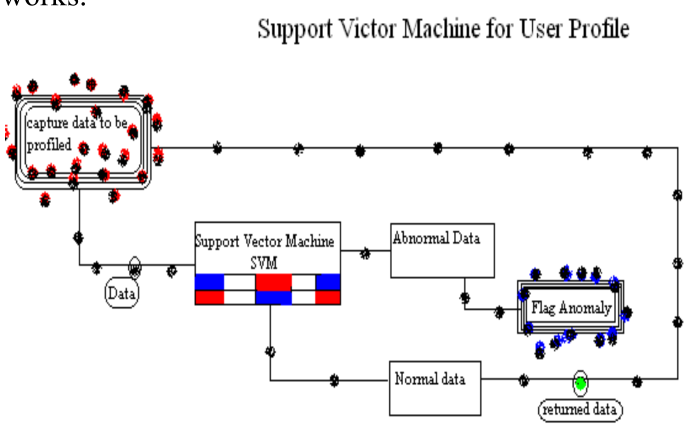


Figure 2, Support vector machine and the user profile filter in the packet filtering

6 *Dataset and Testing Procedure*

The Network security systems have unique testing requirements especially for network analyzer. Like other systems, they need to be tested to ensure that they perform as expected, and to specify the conditions under which they might fail. However, un-like other systems, the data required to perform such testing is not easily or publicly available. So, testing the effectiveness of these types of systems with respect to a given network environment is nearly impossible given the absence of benchmark data sets or testing standards. So, it is not possible or easy to compare the performance, accuracy or efficiency of two systems within a particular type of environment. The most recent attacks should be injected into the dataset that will be used to train such a system to make sure from its ability for identifying recent attacks and diagnosing network problems. After the module has been trained, the system can expose and tested under any dataset such as DARPA 98-99, Lincoln Labs data, or any other private dataset including the real environment. According to system design and other techniques that we are planning to compare this analyzer with, better available dataset which is DARPA99 [6] and real environment. Author will use SVM to classify the dataset to be suitable with the requirements of SVM margin to alert users for attack and filtering process will be done using Flexible Packet Filtering. To compare the results with other software experiments we have to assign specific values that have been got from the analysis for each technique that about to be compared with each other. Also have to consider the environments and the dataset that the software is utilizing for testing their experiments under same circumstances. As it's shown in figure 3, from the analysis of different network monitoring techniques as we assumed that the maximum data captured per mint 40,000 kb\m. and the experiments have showed self similarity with other techniques by using the same dataset and the ability of running the software with any available environment.
The result shows that the Flexible Packet Filtering technique have captured more data in comparing to Traffic Analysis and Monitoring [16], and the hybrid method that is also proposed for anomaly detection by combining both techniques that are Entropy and Support Vector Machine (EaSVM)[2, 30]. That's indicating to the speed of data processing and data filtering. Entropy and Support Vector Machine method (EaSVM) Firstly, they calculate the values of the normalized entropy for different network features. Then SVM model is trained in order to classify the normal traffic vs. attack traffic. In Traffic Analysis and Monitoring, traffic packets are projected to four matrices according to different bytes of the IP address, and then an abnormality detection method for large scale network is proposed. The structure of addresses contained in IPv4 traffic with different length of prefixes is also analyzed. It's important to mention that the targeted environments activities make the traffic more complicated and that would change the network analyzer behavior to produce

better performance and discovering the traffic behavior.

6.1 *System test and Experiment results*

Many researchers has proposed different techniques to capture and recognize anomalies, some uses the IDS and the testing has been done using the DARPA 99 dataset and others used the standard CISCO net-flow analyzer with their own dataset. Most of the network analyzer depends on real dataset and work on real environments. So, in these cases the testing is effected by the environment that the application is running with. The first test has been done in the faculty of computer science (UPM), using the standard method that have followed earlier to develop the software such as method used in Traffic Analyzers and Monitoring [16], and have captured a hug number of traffics with a proportion of 8.2% of error received, these errors might be came by misclassified traffics or it could be unknown attackers or private traffics (Blocked-Traffic) that the network analyzer has no authentication to filter.

The latest test has been done in the same environment but this time using the proposed new method that filter traffics using flexible packet filtering and separating the captured traffics based on their source to enhance the classification operation and using the margin of support vector machine algorithm to classify traffics into normal behavior and anomaly behavior and constructing a reliable user profile, the results has compared with Traffic Analysis and Monitoring (TAaM), and the hybrid method that is also proposed for anomaly detection by combining both techniques that are Entropy and Support Vector Machine (EaSVM)[2, 30]. Our system results appear as the following; almost all types of traffics have been captured and the traffic filtering speed over the number of captured traffics has been increased up to 15% per mint compared to previous methods that are used in majority of nowadays network analyzers. The identification of the network traffic protocols those are traversing over the network have been improved and the result has comes with zero percent of misclassified traffic as shown in figure 4 except for those who are blocked by their private network admin.

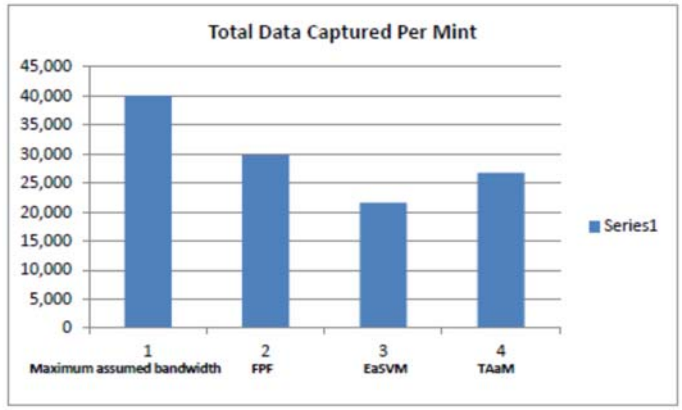


Figure 3 Comparison of total bandwidth captured per mint for each of FPF, EaSVM, and TAaM

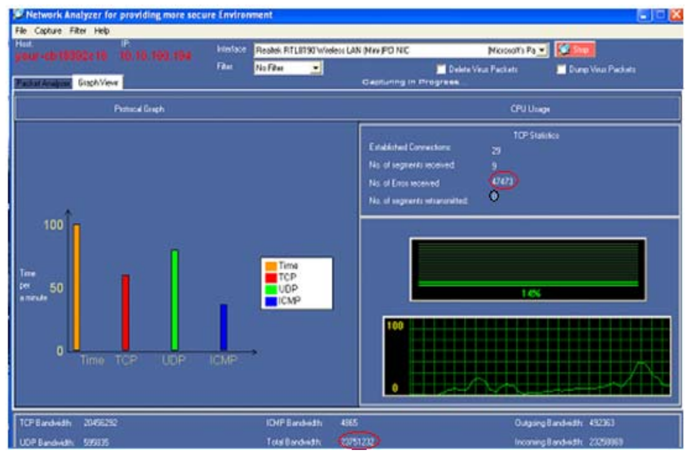


Figure 4 protocols bandwidth rate per mint using the proposed method

Also the traffic signature matching which is known also by misuse detection, gives an excellent results by alerting users to the presence of viruses and attackers those are typically generate a recognizable pattern or “signature” of packets. Figure 4 shows the captured traffics rate per mint with the total bandwidth of each protocol presented in the system interface. And this test has been repeated using DARPA99 dataset and the results compared with TAaM and EaSVM. Table 2 presents the comparison of the results between the proposed Flexible Packet Filtering with TAaM and EaSVM. The results have been gotten by running out the three methods under same environment and dataset. The measurements have been done using our technique which can be easily figured out by looking at the graphic window in figure 4 and monitor the number of error received and total of packet retransmitted which indicating to suspected packets.

Method used to analyze traffics	Total of captured Bandwidth / Mint	Number of errors received / Mint	Dataset used for Experiments	Percentage of Overall Results
FPFaSVM	29,780 Kb/M	0 to 148.9 Kb/M	DARPA99 dataset and real environment	0- 0.5% of errors detected
TAaM	26,732 Kb/M	0 to 614.836 Kb/M	DARPA99 dataset and real environment	0- 2.3% of errors detected
EaSVM	21,601 Kb/M	0 to 259.212 Kb/M	DARPA99 dataset	0- 1.2% of errors detected

Table 2 shows results comparison for FPF, EaSVM and TAaM

First test has been done using the hybrid method that is also proposed for anomaly detection by combining both techniques that are entropy and support vector machine to capture and analyze the data traffic such as EaSVM and second test was measured using the proposed method that is also depends on SVM to classify the captured data. The results shows that by using flexible packet filtering (FPF) and classifying the captured network traffics using the SVM and examining the traffic signature the total of error received has been reduced to 0.5%, whereas; the total of error received using the other methods which is not rely on support vector machine to classify the traffics reach's to almost 3.2% and above, and the total of error received using EaSVM reached to 1.2% of total packet error, and the test of both methods has been done under same circumstances of environment and running time. As we know that total of error received that indicating to suspected traffics or misclassified packets must be as lower as possible to give more reliable information about the captured packets and the calculation can be done by dividing the total of error received over total bandwidth captured.

7 *Conclusion*

In this paper we have merged the analyzed results for both of the flexible packet filtering and support vector machine to get best classification for the captured traffics and to detect anomalies. The purpose is to save time, employees and effort of monitoring the traffics and handling the alarm that indicates the presence of attack. So, we have concluded that by using the SVM alone to classify and detect anomalies traffics do not give very good results as network features are used for learning without processing. But by using the network traffic prediction technique to analyze and detect an anomaly behavior and by applying the Flexible Packet Filtering that will be supported by Traffic Signature Matching including the Traffic source separation technique, the result shows that SVM works perfectly and gives better results than working alone with a proportion of 0.5% of misclassified traffics with the lowest false alarm rate which not exceeds 1% of total filtered traffics. Using the User profile filter our system will be able to identify and detect anomaly behaviors and trace them back to their original source, the tracing is one of the major problems that we should consider in our future work. So, this network analyzer system will be able to detect and identify all types of network attacks and an alert will be triggered indicating to their occurrence, after that it will classify the captured traffics into source, destination, port number and the protocols used to send them over the network.

Acknowledgement

I would like to acknowledge and appreciate the full support that has been given to me through **U**niversity **P**utra **M**alaysia and especially my supervisor Dr. Azizol Abdullah

Author's Contributions

Dr. Mohammed Nazeh Abdulwahid: Participated in all experiments, coordinated the data-analysis and contributed to the writing of the manuscript.
Dr. Azizol Abdullah: Participated in results testing and data-analysis.

Dr. Zuriati Ahmad Zukarnain and **Dr. Nur Izura Udzir** have participated in data analysis and theory vision.

References

1. Barros AK, Vigario R, Jousmaki V, Ohnishi N. Extraction of event-related signals from multi-channel bioelectrical measurements. IEEE Transactions on Biomedical Engineering 2000;47(5):583–8.
2. Basant Agarwal, Namita Mittal, Hybrid Approach for Detection of Anomaly Network Traffic using Data Mining Techniques, Procedia Technology 6 (2012) 996 – 1003.
3. C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics). Springer (2006).
4. Cisco NetFlow. At www.cisco.com/warp/public/732/Tech/netflow/.
5. Cormode G, Muthukrishnan S. An improved data stream summary: the count-min sketch and its applications. In: Proceedings of Latin American theoretical informatics, Buenos Aires, Argentina, 2004. p. 29–38.
6. DAPRAdataset,/http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/docs/detections_1999.htmlS.
7. Duffield N, Lund C, Thorup M. Estimating flow distributions from sampled flow statistics. In: Proceedings of the ACM SIGCOMM, Karlsruhe, Germany, 2003. p. 325–36.
8. E. Al-Shaer, H. Hamed, Firewall policy advisor for anomaly detection and rule editing, in: Mohammed Proc. IEEE/IFIP 8th Int. Symp. Integrated Network Management, IM 2003, March 2003, pp. 17–30.
9. G. Nychis et al.,'" An imperial evaluation of entropy based-anomaly detection". Proceedings of the 8th ACM 2008 SIGCOMM conference on Internet measurement,, ACM Press,, pp 151-156.
10. Giorgi G, Narduzzi C. Detection of anomalous behaviors in networks from traffic measurements. IEEE Transactions on Instrumentation and Measurement 2008;57(12):2782–91.
11. H. Bunke, J. Csirik, Parametric string edit distance and its application to pattern recognition, IEEE Transactions on Systems, Man and Cybernetics 25 (1995) 202.
12. H. Lewis, C. Papadimitriou, Elements of the Theory of Computation, 2nd edition, Prentice Hall, New York,1997.
13. J. Shawe-Taylor, N. Cristianini, Kernel Methods for Pattern Analysis. Cambridge University Press, New York, NY, USA (2004).
14. K. Begnum, M. Burgess, A scaled, immunological approach to anomaly countermeasures, in: Proceedings of the VIII IFIP/IEEE IM Conference on Network Management, 2003.
15. Kim S, Reddy A. A study of analyzing network traffic as images in real-time. In: Proceedings of the IEEE INFOCOM, Miami, USA, 2005. p. 2056–67.
16. Lakhina A, Crovella M, Diot C. Diagnosing network-wide traffic anomalies. In: ACM SIGCOMM; 2004.
17. Lee , Dong Xiang, May 14-16, 2001, Information-Theoretic Measures for Anomaly Detection , Proceedings of the 2001 IEEE Symposium on Security and Privacy, p.130.
18. Lee W, Xiang D. Information-theoretic measures for anomaly detection. In: Proceedings of the IEEE symposium on security and privacy, Oakland, USA, 2001. p. 130–43.
19. Liu H, Feng W, Huang Y, Li X. A peer-to-peer traffic identification method using machine learning. In: Proceedings of the networking, architecture, and storage, Guilin, China, 2007. p. 155–60.
20. Liu Y, Towsley D, Ye T, Bolot J. An information-theoretic approach to network monitoring and measurement. In: Proceedings of the 5th ACM SIGCOMM internet measurement conference, Berkeley, USA, 2005. p. 1–14.
21. Lundin E., Jonsson E., Survey of intrusion detection research, Technical Report, No.02-04, Department of Computer Engineering, Chalmers University of Technology, Goteborg, 2002.
22. M. Roesch, Snort, Intrusion Detection System, http://www.snort.org.
23. M.J. Ranum et al., Implementing a generalized tool for network monitoring, in: Proceedings of the Eleventh Systems Administration Conference (LISA XI), USENIX Association, Berkeley, CA, 1997, p. 1.
24. Marcus Ranum, "False Positives: A User's Guide to Making Sense of IDS Alarms," *white paper*, ICSA labs IDSC, February 2003.
25. N. Damianou, N. Dulay, E.C. Lupu, M. Sloman, and Ponder: a language for specifying security and management policies for distributed systems, Imperial College Research Report Do C 2000/1, 2000.
26. O. Hunaidi, J.H. Rainer, G. Pernica, Measurement and analysis of traffic-induced vibrations, in: Proceedings of the 2nd International Symposium on Transport Noise and Vibrations, St. Petersburg, Russia, 1994, pp. 103–108.

27. P. Barford, J. Kline, D. Plonka, A. Ron, A signal analysis of network traffic anomalies, 2002.
28. P. Hoogenboom, J. Lepreau, Computer system performance problem detection using time series models, in: Proceedings of the USENIX Technical Conference, USENIX Association, Berkeley, CA, 1993, p. 15.
29. Qin T, Guna X, Li W, Wang P. Monitoring abnormal traffic flows based on independent component analysis. In: Proceedings of the international conference on communications, Dresden, Germany, 2009. p. 1–5.
30. R. C. Chen and S. P. Chen, 2008” Intrusion Detection Using a Hybrid Support Vector Machine Based on Entropy and TF-IDF. // international journal of innovative computing information and control (IJICIC), Vol. 4, no2, pp.413-424.
31. R. Durbin, S. Eddy, A. Krigh, G. Mitcheson, Biological Sequence Analysis, Cambridge University Press, Cambridge, 1998.
32. Siris VA, Papagalou F. Application of anomaly detection algorithms for detecting SYN flooding attacks. Global Telecommunication Conf 2004b;29(3):2050e4.
33. Somayaji, S. Forrest, Automated response using system-call delays, in: Proceedings of the 9th USENIX Security Symposium, 2000, p. 185.
34. TCP Dump/libpcap ‘‘TCPDUMP Public Repository’’, http://www.tcpdump.org/; 2009.
35. Thottan M, Ji C. Anomaly detection in IP networks,. IEEE Transactions on Signal Processing 2003; 51(8):2191–204.
36. Ye N, Vilbert S, Chen Q. Computer intrusion detection through EWMA for auto correlated and uncorrelated data. IEEE Transactions on Reliability 2003; 52(1):75–82.
37. Zhang Y, Ge Z, Greenberg A, Roughan M. Network anemography. In: Proceedings of the 5th ACM SIGCOMM conference on Internet measurement, Berkeley, CA, USA, 2005. p. 317–30.
38. Zou C, Gong W, Towsley D, Gao L. The monitoring and early detection of internet worms. IEEE/ACM Transactions on Networking 2005;13(5):961–74.

Corresponding Author: Mohammed N. Abdul Wahid Department of Communication Technology and Networks, Faculty of Computer Science and Information Technology, University Putra Malaysia, Malaysia

Proxy Blind Signcrypion Based on Elliptic Curve

Anwar Sadat*[a], Insaf Ullah[b], Hizbullah Khattak[c], Sultan Ullah [d], Amjad-ur-Rehman[e]

[a]*Department of Information Technology,kohat university of science and Technology K-P, Pakistan*

[b,c,d,e] *Department of Information Technology, Hazara Uuniversity Mansehra K-P, Pakistan*

Abstract

Nowadays anonymity, rights delegations and hiding information play primary role in communications through internet. We proposed a proxy blind signcryption scheme based on elliptic curve discrete logarithm problem (ECDLP) meet all the above requirements. The design scheme is efficient and secure because of elliptic curve crypto system. It meets the security requirements like confidentiality, Message Integrity, Sender public verifiability, Warrant unforgeability, Message Unforgeability, Message Authentication, Proxy Non-Repudiation and blindness. The proposed scheme is best suitable for the devices used in constrained environment.

Keywords: proxy signature, blind signature, elliptic curve, proxy blind signcryption.

1. Introduction

Modern people's needs superfluous effort in less time without showing its identity. In current days anonymity and rights delegation play essential role in promising internet application like digital cash transaction system. It also helps to convert more computation from low resource devices to high resource devices. To protect anonymity of sender Chum [1] introduced blind signature scheme in which the message content is conceal from the signer and signer signing a message blindly. Blind signcryption is the extended version of blind signature which combines both the functionality of blind signature and encryption in single step.

Now to delegates rights Gamage [2] extend the concept of proxy signature into proxy signcryption. It allows the sender to delegates the privileges of signing to proxy and proxy signcrypt message on behalf of the sender. In this paper we introduced a proxy blind signcryption scheme based on elliptic curve discrete logarithm problem (ECDLP).The design scheme enable the original signer to delegates there signing ability to proxy signer and the proxy signer sign a message blindly and send to verifier.

The proposed scheme is design for low resource devices such as pager, smart card and mobile phone due to use elliptic curve shorter size of key.

1.1. Preliminaries

Suppose $p > 3$ is said to be prime number and E is an Elliptic Curve which is defined in equation (1) over finite field F_p :

$$y^2 = x^3 + ax + b \quad (1)$$

Where $a, b \in F_p$ & $4a^3 + 27b^2 \not\equiv 0\ (modp)$. set $E(F_p)$ contains all points $(x, y) \epsilon F_p$ which satisfy the equation(1), with point$\mathcal{O}$ called the point at infinity.

Suppose P& G be the points on elliptic curve$E$, now to a unique integer k from equation $P = k . G$ is called elliptic curve discrete logarithm problem (ECDLP).

The paper is structured as follows. In section 2 we discuss the related work. In section 3 proposed scheme have been discussed. Section 4 discusses the security analysis .section 5 discusses the last conclusion.

2. Related work

Mambo et al. [3] first contribute proxy signature. It enables the sender of a message to give there signing capacity to proxy signer and he signs on behalf of him.

Lin and Jan. [4] first proposed the proxy blind signature scheme. The proposed scheme combined both the functionality of blind and proxy signatures.

Wang et al. [5] contribute a proxy blind signature scheme. The security of a scheme is based on elliptic curve discrete logarithm problem (ECDLP). It does not provide the security properties like strong unforgeability, non-repudiation and unlink ability.

Yang et al. [6] proved Wang et al. proposed an improved proxy blind signature scheme .the scheme is suffers from the original signer's forgery attack and the universal forgery attack.

Qi and Wang. [7] Contribute proxy blind signature scheme .the scheme is based on the hardness of Factoring and ECDLP.it does not meet the properties like unforgeability and unlink ability.

Alghazzawi et al. [8] Proposed proxy blind signature scheme based on elliptic curve discrete logarithm problem. It is insecure against Link ability attacks.

Y. Zheng [9] contributes a new scheme called signcryption which combine the properties of signature and encryption in single step. The security of the proposed signcryption scheme is based on discrete logarithm problem. The scheme ensures the properties like confidentiality and authentication. The cost of signcryption is lesser then the existing signature then encryption scheme. It cannot provide the property of public verifiability and forward security.

In 2009 H. Elkamchouchi et al [10] design a new proxy signcryption scheme based on combination of hard problems like IF,DLP and DHP. They claimed the scheme provide strong security because of these hard problems. But the scheme is not public verifiable and forward secure.In 2013 M. Elkamchouchi et al [11] introduce scheme based on elliptic curve discrete logarithm problem. The claimed the scheme meet all the security requirements.it not public verifiable. Awasthi and Lal [12] design a blind signcryption scheme based on discrete logarithm problem. The design scheme both blind signature and encryption in single step. The limitation of a design scheme is that it not meets the property of public verifiability. Xiuying and Dake [13] design blind signcryption scheme based on discrete logarithm problem which realize the property of public verifiability.

Since there is no proxy blind signcryption scheme based on ECDLP is available in literature. In this paper we proposed a proxy blind signcryption scheme based on elliptic curve discrete logarithm problem.

3. Proposed scheme

In this section we present our proposed proxy blind signcrypion scheme based on elliptic curve discrete logarithm problem. Proposed scheme contain the following phases.

3.1. Notations

H/kh: is an irreversible hash functions and keyed hash function

Q: Is an huge prime number ,Q $\geq 2^{60}$

FQ:As in finite field having order Q

E_{K_1}:Encryption through symmetric key

D_{k1}:Decryption through symmetric key

E: Secure elliptic curve $E: B^2 = A^3 + xA + y \ mod \ Q$

X, y : be the two integers $,(x, y) < Q \ and \ (4x2 + 27y2) \ mod \ Q$

Mesg: plaintext/message

Cip: Cipher text/encrypted message

3.2. Key Generation

Table 1 shows the generations of key pairs of Sender, Proxy and verifier/recipient as:

Table 1: Key Generation

u_a	Private key sender , $u_a \in \{0, 1, 2, \ldots\ldots, q-1\}$
v_a	Public key of Sender, $v_a = u_a G$.
u_p	Proxy Private key , $u_p \in \{0, 1, 2, \ldots\ldots, q-1\}$
v_p	Public key of proxy, $v_p = u_p G$.
u_s	Private of signer, $x_s \in \{0, 1, 2, \ldots\ldots, q-1\}$
v_s	Public key of signer, $Y_s = x_s G$
u_r	Private key of verifier,$x_v \in \{0, 1, 2, \ldots\ldots, q-1\}$
v_r	Public key of verifier,$Y_v = x_v G$.

3.3. Proxy key generation

1. Randomly chose l
2. Calculate $\varphi = l.G$
3. Calculate $\text{Ɣ} = \left(l - u_a.h(\varphi, m_w)\right) mod q$

   Send $(\varphi, \text{Ɣ}, m_w)$ to proxy

3.4. Proxy verification

1. Compute $\varphi' = \text{Ɣ}.G + h(\varphi, m_w).v_a$

3.5. Proxy

(1) Randomly choose blinding factorsδ , μ and $\Omega \in \{0,1,2, \ldots \ldots . \mathrm{n}-1\}$

(2) Calculate $K = (\mathrm{K}_1 \parallel \mathrm{K}_2) = \Omega.\mathrm{y}_\mathrm{v} \bmod \mathrm{n}$

(3) Split K= $(\mathrm{K}_1, \mathrm{K}_2)$

(4) Calculate $\lambda = \mathrm{h}\ (\mathrm{msge}\ \parallel \mathrm{K}_2)$

(5) Calculate $\mathrm{c} = \mathrm{E}_{\mathrm{K}_1}(\mathrm{msge})$

(6) $J = ((\Omega + \mu).\Upsilon + \delta.G) mod n$

(7) $\varpi = (\lambda + \mu) mod\ n$

(8) Send ϖ to signer

3.6. Signer

(1) calculate $\bar{S} = (x_s + \varpi.w) mod\ n$

(2) Send $\bar{S}$to proxy

3.7. Proxy

(1) calculate $S = \frac{\Omega}{\lambda + \bar{S} + \delta} modn$

(2) Send (λ, S, J) to Bob.

3.8. Unsigncryption

(1) calculate $\chi = u_r.S$

(2) calculate $k = \chi.(y_s + J + \lambda.G)$

(3) calculate $m = D_{k1}(c)$

(4) calculate $\lambda' = h(m \parallel K_2)$

(5) Accept m as a valid original message if $\lambda' = \lambda$ otherwise reject

4. Correctness Analysis

Proof 01:

Proxy signer checks the validity by using the following:

$$\gamma.G + h(\varphi, m_w).v_a$$

$$= \left(l - u_a.h(\varphi, m_w)\right).G + h(\varphi, m_w).v_a$$

$$= \left(l - u_a.h(\varphi, m_w)\right).G + h(\varphi, m_w).u_a.G$$

$$= G(\left(l - u_a.h(\varphi, m_w)\right). + h(\varphi, m_w).u_a)$$

$$= G(l - u_a.h(\varphi, m_w) + u_a.h(\varphi, m_w))$$

$$= G(l - u_a.h(\varphi, m_w) + u_a.h(\varphi, m_w))$$

$$= l.G = \varphi$$

Proof 02: The validity proof of scheme is shown below.

$$k = (y_s + T + r.G)\, x_v.s$$

$$= x_v.s\,(x_s.G + (r + \beta).z + \alpha.G + r.G)$$

$$= x_v.s\,(x_s.G + r.z + \beta.z + \alpha.G + r.G)$$

$$= x_v.s\,(x_s.G + r.G + \alpha.G + r.z + \beta.z)$$

$$= \frac{\gamma x_v((G(x_s + r + \alpha) + z(r + \beta))}{(r + \bar{s} + \alpha)}$$

$$= \frac{\gamma x_v(G(x_s + r + \alpha) + w.G(r + \beta))}{(r + \alpha + x_s + \bar{r}.w)}$$

$$= \frac{\gamma x_v G(x_s + r + \alpha + w(r + \beta))}{(r + \alpha + x_s + w(r + \beta))}$$

$$= \gamma x_v G$$

$= \gamma\, y_v$ *which is true.*

5. Security Analysis

In this section we discuss the security requirements of a proposed scheme such as confidentiality, message integrity, Sender public verifiability, Warrant unforgeability, Message Unforgeability, Message Authentication, Proxy Non-Repudiation and Blindness.

5.1. Confidentiality

The proposed proxy blind signcrypion scheme provides the property of message confidentiality. When the attacker try to reveal the contents of message then it must get the secret key k from $k = (\Omega.y_v)$.thus it is hard and equivalent to solve elliptic curve discrete logarithm problem (ECDLP) for eavesdropper. The attacker can also get easily Ω from $\mathrm{S} = \frac{\Omega}{\lambda+\bar{\mathrm{S}}+\delta}$ but it is computationally hard and equal to finding two unknown variables from one equation.

5.2. Message Integrity

Our proposed scheme uses one way hash function to provide integrity. When the eavesdropper try to covert cipher text c into ć, then the message m will also be convert toḿ. But one way hash function meet the property of collision resistant $\acute{\mathrm{R}} = \mathrm{h}\,(\acute{\mathrm{m}} \parallel \mathrm{K}_2) \neq \mathrm{R} = \mathrm{h}\,(\mathrm{m} \parallel \mathrm{K}_2)$ so, change in C can be detected easily.

5.3. Sender public verifiability

Our design scheme provides the property of sender public verifiability. Using $\left(Y_a = T - \lambda.\frac{D}{h(\alpha, m_w)}\right)$ anyone can verify the warrant m_w is send by the sender or not.

5.4. Warrant unforgeability

The design scheme also ensures the security property warrant unforgeability. When the attacker tries to compute valid signature then it get d and x_a from $\lambda = \left(d - x_a.h(Z, m_w)\right)$.Therefore, finding to exact variables from same equation is computationally infeasible.

5.5. Message Unforgeability

Our proposed scheme provides the property of message unforgeability. When the attacker tries to compute original signature then he must solve $\bar{\mathrm{S}} = (\mathrm{x_s} + \mathrm{r.d})$. for this the eavesdropper first get $\mathrm{x_a}$ from$\mathrm{Y_s} = \mathrm{x_s}.\mathrm{G}$ which is computationally hard for attacker and equal to solve elliptic curve discrete logarithm problem (ECDLP).also required d from$\mathrm{Z} = \mathrm{d.G}$ is hard to calculate for eavesdropper and equivalent to solve elliptic curve discrete logarithm problem (ECDLP).

5.6. Message Authentication

In our proposed scheme the sender use their own private key to generate signature $\bar{\mathrm{S}} = (\mathrm{x_s} + \mathrm{r.d})$ if the attacker wants to generate a valid signature then it must get a private key $\mathrm{x_s}$ from$\mathrm{Y_s} = \mathrm{x_s}.\mathrm{G}$ which is computationally equivalent to solve elliptic curve discrete logarithm problem (ECDLP).

5.7. Proxy Non-Repudiation

Our proposed scheme provides the property of non-repudiation. In design scheme when dispute occur between sender and receiver then the trusted party use $\mathrm{k_2}$ to calculate $\mathrm{R} = \mathrm{h(m} \parallel \mathrm{K_2)}$ and $\mathrm{R'} = \mathrm{h(\,m} \parallel \mathrm{K_2)}$. If$\mathrm{R} = \mathrm{R'}$ then the signature generated by sender otherwise not.

5.8. Blindness

In our proposed scheme the signer select blind factors to generate blind message. In design scheme the signer cannot know about blind factors and the contents of a message.

6. Conclusion

This paper presents a new idea called proxy blind signcrypion scheme based on elliptic curve discrete logarithm problem. The scheme combined both the properties of proxy and blind signcryption. It also meets the security properties like confidentiality, message integrity, Sender public verifiability, Warrant unforgeability, Message Unforgeability, Message Authentication, Proxy Non-Repudiation and Blindness .The scheme is efficient because of elliptic curve cryptosystem. It is best suitable for low resource devices.

References

[1] D Chaum, "Blind signature for untraceable payments",Advances in Cryptology, proceeding of CRYPTO'82,Springer-Verlag, New York, 1983, pp.199-203.

[2] Gamage et al (1999) an efficient scheme for secure message transmission using proxy-signcryption. In Proceedings of the 22nd Australasian Computer Science Conference, pp. 420–431, Springer.

[3] M. Mambo, K. Usuda, and E. Okamoto, "Proxy signatures: Delegation of the power to sign messages", IEICE Transaction on Fundamentals, E79-A (1996), pp.1338-1353.

[4] W. D. Lin and J. K. Jan, "A security personal learning tools using a proxy blind signature scheme," in Proc. of Intl Conference on Chinese Language Computing, 2000,pp. 273-277.

[5] H. Y. Wang and R. C. Wang, "A proxy blind signature scheme based on ECDLP", Chinese Journal of Electronics, Vol.14, No.2, 2005, pp. 281-284.

[6] Xuan Yang and Zhaoping Yu, "Security Analysis of a Proxy Blind Signature Scheme Based on ECDLP", The 4th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM'08), Oct. 2008, pp. 1-4.

[7] C. Qi and Y. Wang , "An Improved proxy blind signature scheme based on factoring and ECDLP" in Proc. International Conference on Computational Intelligence and Software Engineering, 2009, Dec.11-13, 2009, pp. 1-4.

[8] D. M. Alghazzawi, T. M. Salim and S. H. Hasan, "A new proxy blind signature scheme based on ECDLP IJCSI International Journal of Computer Science Issues, vol. 8, issue 3, no. 1, May 2011.

[9] Y. Zheng. "Digital signcryption or how to achieve cost (signature encryption) << cost (signature) + cost (encryption)" In CRYPTO '97: Proceedings of the 17th Annual International Cryptology Conference on Advances in Cryptology, pp 165-179, London, UK, 1997. Springer-Verlag.

[10] Elkamchouchi et al (2009) a new efficient strong proxy Signcryption scheme based on a combination of hard problems, IEEE International Conference on Systems, Man and Cybernetics.

[11] Elkamchouchi et al (2013) An Efficient Proxy Signcryption Scheme Based on the Discrete Logarithm Problem. *International Journal of Information Technology, Modeling and Computing (IJITMC) Vol* 1.

[12] A. K. Awasthi, Sunder L .An Efficient Scheme for Sensitive Message Transmission Using Blind Signcryption[EB/OL].[2007-08-23].http://arxiv.org/abs/cs.CR/0504095.

[13] YU Xiuying and HE Dake. "A new efficient blind Signcryption scheme": Wuhan university journal of natural sciences; 2008, Vol.13 No.6, 662-664.

A Comprehensive Survey on Hardware/Software Partitioning Process in Co-Design

Imene Mhadhbi[1], Slim BEN OTHMAN[1] and Slim Ben Saoud[1]
[1]*Department of Electrical Engineering, National Institute of Applied Sciences and Technology, Polytechnic School of Tunisia, Advanced Systems Laboratory, B.P. 676, 1080 Tunis Cedex, Tunisia*

***Abstract*-Co-design methodology deals with the problem of designing complex embedded systems, where Hardware/software partitioning is one key challenge. It decides strategically the system's tasks that will be executed on general purpose units and the ones implemented on dedicated hardware units, based on a set of constraints. Many relevant studies and contributions about the automation techniques of the partitioning step exist. In this work, we explore the concept of the hardware/software partitioning process. We also provide an overview about the historical achievements and highlight the future research directions of this co-design process.**

***Keywords*: Co-design; embedded system; hardware/software partitioning; embedded architecture**

I. Introduction

Modern embedded systems are rapidly becoming an important factor of the exponential growth of e-industry due to their progressively sophisticated functionalities. Designers of these modern embedded systems have continually proposed new design methodologies and architectures. Recently, embedded system architectures incorporated both hardware and software components. Traditionally, the design of hardware and software was developed separately in the early stages of the co-design process [1]. Since 1990s, the Co-design methodology has emerged as a new research subject to design complex embedded systems. It has included several tasks such as modeling, hardware/software partitioning, scheduling, validation and implementation [2, 3]. One of the most important tasks in the Co-design is the hardware/software partitioning process. It can be defined as a cooperative design of hardware and software tasks to achieve the best performances of the designed embedded system. The hardware/software partitioning process was carried out manually [4]. The manually decision was limited to small design problems with small number of components. With the increasing of embedded applications and architectures complexities, automatic partitioning has emerged as a NP-Hard problem [5-7].

Different methods and techniques are applied to automate the partitioning process. In this paper, we will attempt to examine the importance of this co-design process. We will review the different partitioning techniques and approaches on a respectable volume of references. The rest of the paper is organized as follow; the next section gives an overview of the co-design methodology. Section 3 outlines the partitioning process from initial specification to target implementation. The studying of the different partitioning techniques and approaches are presented, respectively, in the sections 3, 4 and 5. Finally, section 6 concludes the paper.

II. Hardware/Software Co-Design Approach

The co-design methodology was appeared as a new design methodology to design complex embedded systems [8]. It presents an important issue to increase the performances of the designed embedded system. The co-design methodology includes several subtasks [2, 3]. These subtasks reside on modeling, partitioning, scheduling, validation and implementation.

i. The modeling process is the step of defining the system specifications. This specification can directly affect the designed embedded system performances [9]. Some researches concentrated on raising the design level of the specification in order to speed up the implementation of complex embedded systems, rise their performances and expand the time reserved to the optimization and the final circuits refinements using High Level Synthesis (HLS) tools [10].

ii. The Partitioning and scheduling processes present the step through which the embedded application is partitioned and then scheduled. The partitioning represents the step of deciding which tasks are able to be to be executed on software units and which ones implemented on hardware cores. However, the scheduling represents the step of organizing the set of tasks based on deadlines. Diverse studies propose to automate these steps in order to attain better performances. Some attempt to improve the partitioning and the scheduling processes together [11-13], while others focus on exclusively partitioning step.

iii. The validation process attempts to prove that the embedded system works as designed. Several techniques have been presented. The Co-simulation process has been proposed [8] in order to speed up the system design when executing simulation of heterogeneous systems whose hardware and software architecture interacts. the Hardware in the loop (HIL) technique is also proposed [14] to support simulation and validation of the heterogeneous hardware/software system co-design.

iv. The implementing process presents the step of the physical implementation of the hardware tasks (through synthesis) and executable software tasks (through compilation). Different propositions have been presented in this implementation step. Several of them have been focused on minimizing the implementation process by transforming the behavioral description of the embedded systems into fully structural netlist system components using a high-level input language such as SpecC or Bluespec [15, 16].

Significant researches have been developed, in recent years, to improve the co-design methodology from the input specification to the system's validation and implementation. It can be stated that hardware/software partitioning process is one of the main phases during co-design. Different studies highlighted the automation of this process to increase the embedded systems performances. In the next section, we will present a contextual state of art and related works on the hardware/software partitioning step of the co-design methodology.

II. Hardware/Software Partitioning Process

The Hardware/software partitioning presents a crucial step in the co-design process. It has the major impact on the cost/performance characteristics of the designed system. Figure 1 gives an overview of the basic flow of the partitioning process.

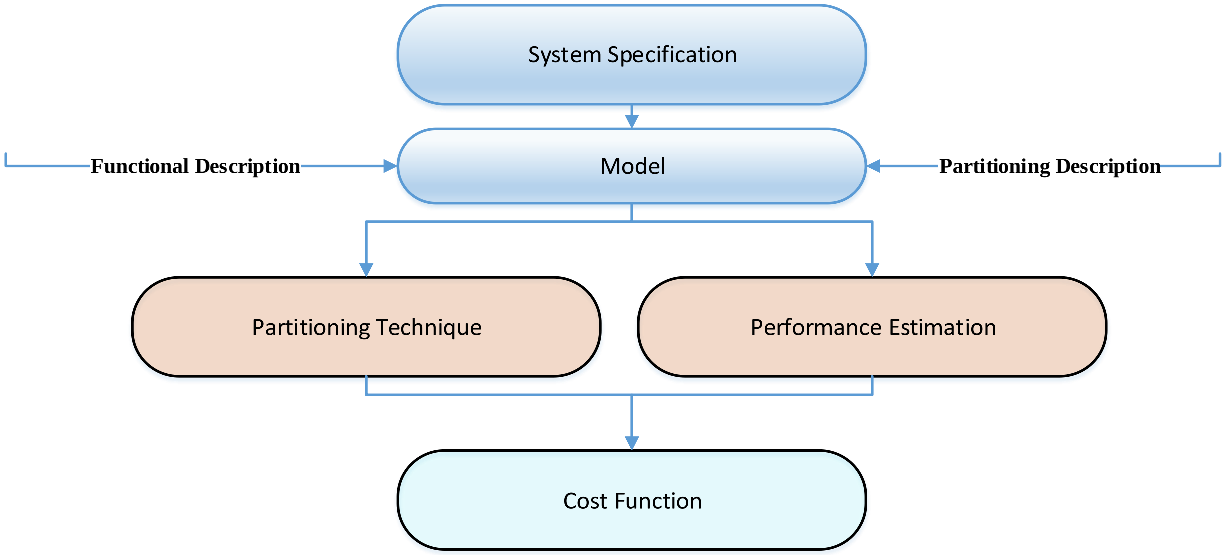


Figure 1. Basic Flow of the partitioning process.

According to Hidalgo et al. [17], the partitioning can be classified into: structural and functional partitioning process. In the structural partitioning, the embedded system is firstly synthesized and then partitioned into blocks. This partitioning way is very popular. However, it is difficult to make corrections into design and usually the number of blocks is very high using the structural partitioning. In the other hand, in the functional partitioning, tasks are divided into multiple sub-specifications. Each sub-specification denotes the functionality of a system component such as a custom-hardware or software processor. It is compiled down to assembler code or synthesized down to gates. The functional partitioning process has numerous advantages that make a most used way for hardware/software partitioning. Figure 2 illustrates the implementation method of structural and functional partitioning implementation.

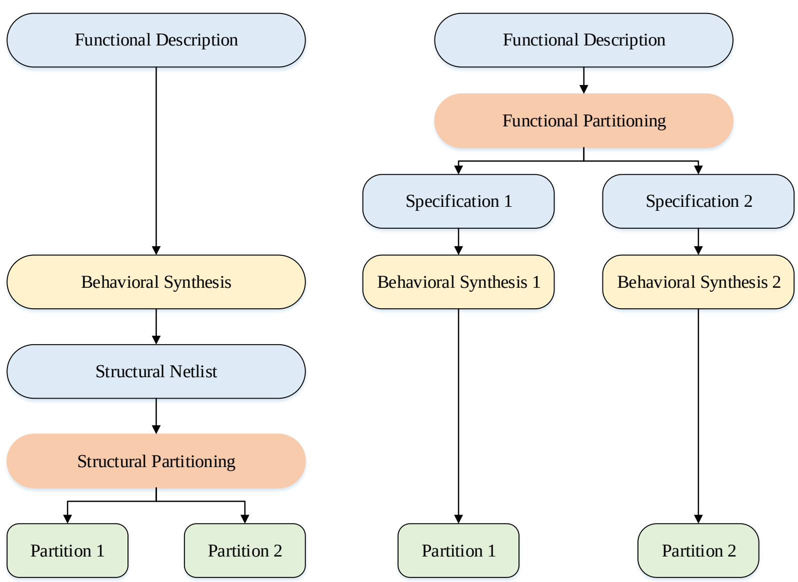


Figure 2. Structural partitioning and Function partitioning processes.

In this paper, we will expose the different parameters of partitioning strategy from the initial system specification to the final partitioning decision. These parameters are: the system specification, the performance estimation, the cost function and the target architectures.

A. System Specification: models and granularities

Before starting the partitioning process, it is necessary to transform the initial specification to the formal specification (model). The choice of formalization has an impact on the quality of the system design. The formal description offer different levels of abstraction called granularity. The granularity determines how to divide an application into a set of n fragments B, where: B = {b0, b1,… bn}, in order to affect these fragments to software/hardware components of the target architecture. Different levels of granularity exist. We note: (i) the coarse granularity: the level of abstraction is presented as an object or entity, (ii) the middle granularity: the level of abstraction is presented as functions, procedures or processes, and (iii) finally, the fine granularity: the level of abstraction is presented as single or arithmetic operation instruction.

The partitioning of the embedded application specification based on coarse and intermediate granularities can be performed manually. However, the partitioning based on fine granularity is more complex due to huge number of entities to be mapped.

The modeling of embedded applications presents a complex process because of their heterogeneity. Many modeling approaches have been used for the hardware/software co-design methodology. Several computation models have been developed and used to represent heterogeneous embedded applications such as Finite State Machines (FSM), Petri Nets, Data Flow Graph (DFG) [18-20], Control Flow Graph (CFG) [21], Control Data Flow Graph (CDFG) [22], Direct Acyclic Graph (DAG), State Transition Graph (STG) [23], Synchronous/Reactive Models, Communicating Processes, etc.

The most used computational model in the partitioning problem has always been the DAG graph. A node, in a DAG graph, represents a task which is a set of instructions that must be executed sequentially in the same unit without preemption. The DAG graphs can be randomly generated using TGFF tool [24] with a random or uniform distribution. The uniform distribution is generally presented as in-tree, out-tree [25], fork-joint [26], mean valued analysis and FFT [27]. According to [5, 19], the topology of the input specification can affect partitioning performances, to a certain extent. In [9], the author compares the improvement of a proposed partitioning approach over random and uniform DAGs graphs. The applied uniform graphs are the FFT, the Fork-joint, the in-tree and the out-tree graphs. Results prove that the uniform graphs generate between 5% to 80% better solutions than random graph. However, in [28], the authors focus on the best partitioning technique, using different input specifications such as random graph, out-tree, in-tree, fork-joint, mean value and FFT graphs. Results prove that the best performance is provided from the random and the out-tree graphs.

After the application's modeling, it is necessary to assign to each node the related cost that are obtained after a performance estimation procedure.

B. Performance Estimation Procedure

The performance estimation tools generate the necessary information about the performance of application's entities in order to collect, analyze and calculate the estimated costs throughout the co-design cycle. The Profiling is the most popular tool for estimating software costs. However, the hardware costs are estimated using different tools. The Xilinx System Generator tool, the Xilinx ISE Analyzer and the Timer estimates hardware performance estimated the hardware resources utilization. However, the analyzer timer estimates both area and power performance. The performance estimation for each node is necessary to compute the cost function [29], as will be discussed in the next sub-section.

C. The Cost Function

The mail subject of the software/hardware partitioning process is to get the best mapping of the application entities on hardware/software components based on constraints. These constraints are usually related to: (i) the performance aspect which includes the software/hardware processing time, the latency, the power consumption, etc., (ii) the manufacturing aspect which comprises the silicon surface, the number of logic gates or transistors for the hardware implementation, the number of words occupied by the data and code for the software implementation, etc. (iii) the violation aspect which presents the limits of material resources available, the maximum memory available, the maximum power, etc., (vi) the communication aspect which characterizes the communication data between hardware and software components, etc. In order to find the best partitioning, the partitioning decision is based on the computation of all constraints in the same function called cost function (or objective function).

The growing complexity of embedded applications recommend the consideration of multiple constraints in the objective function which is transformed from a simple to a multi-objective function. In [30], the authors study a compromise between two objectives metrics: buffer size and system delay. In [31], the authors consider also different constraints including hardware resources, execution time and implementation styles of architecture (pipeline, multi-cycle operation, etc.). In [22], the authors consider a bi-objective partitioning problem including a six pair of combination terms between execution time, slice rate, memory requirement, and power consumption. Results prove that it is sufficient to study partitioning problem using memory requirement and slice rate terms by taking into account the conflicting nature of cost metrics. Some studies are based on different objective functions to find the best partitioning solution. In [32], the authors suggest four objective functions based on the time, the power, the time and power and the resources allocation constraints. In [28], the authors apply two different objective functions to determine which one give better solution. Results prove that the first function which involves the minimum area parameter provides a solution with shorter execution time. Some other studies also prove that reducing a constraint can affect the performance of the partitioning technique, especially the communication cost. In [24, 33, 34], the authors focus on minimizing the compromise between the hardware area and execution time without any consideration of the communication cost. In [18, 19, 35], the authors consider communication cost as an important issue in the partitioning problem. Indeed, a comparison between two partitioning techniques reported in [36] proves that the first technique yields the best solution comparing to the second one based on minimizing the area, the latency and the execution time without taking into account communication cost. However, in [37], the second technique was achieved better. In this study, the author minimizes the communication cost without taking into account resource conflicts. Furthermore, in [38], the author proves that the second technique can be better applied to the partitioning problem. The proceeding work is based on minimizing the timing constraints and the hardware cost without considering the communication constraints. The hardware/software partitioning requires an efficient cost function and an effective technique.

C. *Hardware/Software Partitioning Techniques*

Different techniques have been developed to automate the hardware/software partitioning process. These techniques can be classified based on the degree of automation into: static, semi-static and dynamic techniques.

- The static partitioning techniques:

The static partitioning techniques are generally based on scenarios taken at worst WCET (Worst Case Execution Time) to ensure the satisfaction of real-time constraints. They are presented as offline techniques and need a preliminary study of the application, architecture and execution environment. Different static techniques have been applied to automate the hardware/software partitioning process. These techniques are based on exact or heuristic algorithms.

Exact algorithms ensure generally the generation of optimal solutions. These algorithms can be divided into general and special classes. General classes present the most used algorithms in the partitioning problem. In this context, we can find branch-and-bound algorithm [39], the branch-and-cut method, the Integer Linear Programming (ILP) algorithm [19, 40] and dynamic programming algorithm [41, 42]. Static algorithms have a several drawbacks: firstly, they are very slow and can be applied only for graphs of small sizes. Second, they are very greedy in memory consumption. Third, they require a huge development time. Finally, they often difficult to be modified if some details of the cost function are changed. To overcome these drawbacks, researchers have migrated to the heuristic algorithms due to their flexibilities and efficiencies.

Heuristic algorithms are proposed to partition of graphs with large number of nodes. These algorithms can be classified according to their structures and their used search process, as described in the Table 1.

TABLE I
CLASSIFICATION OF HEURISTIC ALGORITHMS

Classification of Heuristic algorithms	Classification	Types		Algorithms
Structure classification	Decision making	**Determinists algorithms**: the same initial input always leads to the same final solution		TS, Greedy, Hill climbing, etc.
		Stochastic algorithms: different solutions can be generated from a single input		SA, ACO, GA, PSO, etc.
	Solution space utilization	**Constructive algorithms**: create the candidates by sequentially adding the components of the solution until a complete feasible solution is reached.		Greedy and Hierarchical Clustering, etc.
		Iterative algorithms: perform research in parallel in a space of candidates. The best candidate is selected as the optimal solution.	**Single-point algorithms**	SA, TS, Hill climbing, etc.
			population-based algorithms	GA, Memetic, PSO, ACO, etc.
Search Process	Trajectory type	Direct Trajectory algorithms: represented as a single trajectory (or path) in the representative neighborhood graph.		SA, local search, TS, Lin\Kernighan, etc.
		Discontinuous trajectory algorithms: follow a discontinuous walk with respect of the neighborhood graph		GA, ACO, Greedy, etc.
	Problem model	Instance-based algorithms: generate candidate using solely		SA, Iterated local

	presence	the current candidate as the current population solutions	search, GA, etc.
		Model-based algorithms: candidate are generated using a probabilistic parameter model that is updating using the previous seen candidate	ACO, etc.

Several studies on partitioning have applied static techniques to generate optimal solutions. Indeed, choosing the best suitable partitioning algorithm for a specific application and a predefined target architecture is difficult. Some studies propose to compare heuristic algorithms based on the same partitioning parameters such as: the same constraints, the same architecture and the same target application. We have presented advantages and disadvantages of some algorithms on the Table 2.

TABLE II
SOME HEURISTIC ALGORITHMS: ADVANTAGES AND HANDICAPS

Algorithms	Publication	Objective metrics	Advantages	Handicaps
GA	[43]	Area constraint Processing time	GA is adaptable and effective for solving combinatorial optimization problems	GA demands more memory to store information about a large number of solutions
	[5]	Hardware area Execution time	GA is better than ACO in term of cumulative cost.	GA consumes too much time comparing to ACO, ILP and PSO
PSO	[5]	Hardware area Execution time	PSO outperforms GA and ACO algorithms from the point of views of cumulative runtime and cost.	ILP is better than PSO in term of cumulative running time and cumulative cost.
ACO	[5]	Hardware area Execution time	ACO is better than GA in term of cumulative time	ACO consumes too many resources comparing to GA, PSO and ILP
SA	[43]	Area constraint Processing time	SA avoids becoming trapped in local targets by using Boltzmann distribution	SA are rapidly in both processing time and search time comparing to GA
TS	[43]	Area constraint Processing time	TS systematic approach to searching a solutions space to avoid acyclic searching or being trapped in local targets.	TS presents the shortest processing time and execution time comparing to GA and SA

These comparative studies, although old, may be the source of several suggestions for improvement or proposition of new partitioning techniques.

- The semi-static partitioning techniques:

The semi-static techniques are more recent than the static techniques. In semi-static techniques, the partitioning change decisions are based on the results found using static partitioning technique. They are based on a static study (offline) and a complementary analysis (online). These techniques target real-time constraints applications. They operate on runtime tasks corresponding to the worst cases WCET. Since their appearance, these techniques try to adapt the processing resources to the task needs taking into account the time constraints of all available resources. This semi-static partitioning technique includes these steps: (i) searching all paths of possible executions, (ii) representing the curve of the execution time of the task in accordance with a correlation parameter as a segments and associating the maximum execution time of the segment to the entire segments, (iii) transforming the DFG task graph by duplication of each task as many times as there are segments identified on the curve associated with a task. Finally, (vi) applying a heuristic or exact algorithm on each of these paths to build configurations and load it into the architecture of memory. The semi-static techniques are applied in different partitioning studies [44]. Indeed, the proposed work in [45] seem among the first to use a semi-static technique. The architecture proposed in this work is constituted by a general purpose processor connected to a dynamically reconfigurable unit. Yu Kwang [44] also offers a semi-static partitioning methodology called 'On-Off methodology'. The principle is to use online derived implementations prepared from offline technique based on a genetic algorithm.

- The dynamic partitioning techniques:

In the dynamic techniques, partitioning can change according to the needs of the system design. These techniques allow the system to self-adjust to the application execution environment. The difference between the dynamic partitioning techniques and the semi-static techniques occurs in the absence of processing portions performed offline. Thus dynamic partitioning techniques are able to solve the problem of partitioning online. The target applications by these techniques are those which include variable characteristics as a function of data to be processed.

Several studies partitioning software/hardware are based on these techniques. G. Stittet al. [46] propose a dynamic partitioning technique based on a profiling stage to detect the most critical software in loops runtime. In [47], G. Stittet al. have also proposed the same work more detailed presenting the tools used their proposed dynamic technique namely; the profiling tools, the decompiling tools, the synthesis tools and the placement/routing online tools. The partitioning method proposed follows the following strategy: (1) searching for critical parts for profiling, (2) decompiling software code, (3) behavioral synthesis, (4) logic synthesis, (5) placement and routing, and finally, (6) updating the software for communicating with the hardware. In [48], the authors propose a technique of partitioning based on a charges balancing and an evolutionary heuristics. Depending on the change in the calculation power demand, the system responds with a dynamic load distribution on available resources. A major drawback of this technique is that the distribution changes are not predicted; there are significant data loss during the transitional arrangements. The work presented in [49] also examines the problem of dynamic partitioning. The authors present an architecture adapted for real-time applications with low power consumption.

D. The Target Architecture

Designers of embedded systems preselect the target architecture in the early stage of the design process to reduce the design space. The target architecture presents a description of the number and the type of components proposed to implement the embedded system and the connection between these components. This architecture is usually characterized by including programmable devices (standard processors, microcontrollers, etc.), dedicated devices (ASICs, FPGAs, etc.) and communication components. Partitioning process can be classified based on the target architecture as binary and extended approaches. The main difference between these two kinds of partitioning approaches appears at the number of the used components and their categories.

III. The Binary Partitioning Approach

The Binary partitioning approach presents the problem of mapping an application's tasks (or nodes) into two parts; one part executes as sequential instructions on software component and a second part that runs as parallel circuits on hardware component to achieve the best embedded system performance. In the binary approach, the number of the possible partitioning solution for N tasks is 2N.

Many researchers are committed to apply standards partitioning techniques. In particular, the GA algorithm [22], the Kernighan/Lin algorithm [50], the SA algorithm [51], the TS algorithm [18], the multi-level partitioning (MLP) algorithm [52], the recursive spectral bisection (RSB) algorithm [53], the hardware-oriented partitioning (HOP) algorithm [54], the enhancement partitioning algorithm [55], the efficiently partitioning algorithm [56], the sophisticated computer partitioning algorithm [57], etc. Many other researchers have focused to improve the performance of the binary partition approach by the addition of a parameter or a combination of two binary approaches.

A. Improvement of existent Binary partitioning techniques

Many interesting researches have improved the existing binary partitioning techniques in order to find optimal partitioning solution. Some studies are focused to improve the GA algorithm. In [58], An Advanced Non-Dominated Sorting Genetic Algorithm (ANSGA) was introduced by proposing a removing technique for building Non-Dominated Sorting (NDS) to reduce the computational problem and obtain a good partitioning solution for SoC architecture.

In the same context, in [59], the authors propose a new partitioning technique applied on a system that contains only CPU. They use the hardware orientation technique to create the initial solution of the GA and reduce the crossover and the mutation probability to get good partitioning solution. Experimental results show that the proposed algorithm outperforms GA and ANSGA algorithms and ANSGA algorithms. Furthermore, in [60], an efficient crossover operator, called DSO, is proposed to improve the speed of the algorithm to reach the optimal partitioning solution. The authors use the GA's crossover operator to provide optimal solution for solving the fitness function problem in the GA for partitioning in reconfigurable embedded systems architectures. In [36], the authors propose a new partitioning algorithm, for SoPC architecture, based on the principle of Binary Search Trees (BST) algorithm and GA. Results prove that this algorithm reduces the logic area compared to TS, GA et SA algorithms.

Many studies are focused on improving the PSO algorithm. In [61], the authors suggest a modified PSO restarting technique, named the Re-excited PSO algorithm, to reduce the design size and fix the mapping of design components based on reconfigurable FPGA system. In [62], an effective hybrid multi-objective partitioning algorithm, based on discrete particle swarm optimization (DPSO) with local search strategy, called MDPSO-LS is proposed to solve VLSI two-way partitioning problem.

Many research papers are emphasized to increase the performance of the SA algorithm. In [43], new version of SA algorithm, named Localized Simulated Annealing (LSA) algorithm, was developed based on simple architecture (one hardware processor and one software processor). It divides search space and provides a better control over temperature and annealing speed in different subspaces. In the same context, in [63], the authors improve the disturbance model of the annealing schedule by proposing a new cost function method to accelerate the convergence speed of SA algorithm. The proposed algorithm reduces the running time and increases the probability of finding an optimal solution for simple embedded architecture compared to the classical SA algorithm. In [43], other version of TS, called Tabu Search with Penalty Reward (TSPR), is presented to partition simple architecture. This version offers better results compared to those from standard TS. In the paper [64], the authors intend also, to rise the Hill Climbing algorithm performance by applying fuzzy logic to model the uncertainty of variables involved in the decision criteria. Comparing to Random search, GA, SA, TS, ES and Hill Climbing algorithms, the proposed RHC algorithm generates the best performing solution.

B. Combination between existent Binary partitioning techniques

Designers focus, also to achieve more optimal partitioning solutions by emphasizing a combination between two existing partitioning algorithms.

In [65], the authors proposes a genetic particle swarm optimization (GPSO) algorithm. The combination between GA and PSO algorithms is made in order to get the fastest convergence speed of PSO algorithm and the easy use of the GA in solving partitioning problem for a single CPU embedded systems. In [66], the authors introduced a genetic simulated annealing (GSA) algorithm that combine GA and SA to solve partitioning optimization problem. GA presents a strong search capability while SA algorithm will fail in a local optimal solution easily. They conclude that the combination between these two algorithms provide more accurate solution faster. Experiment results show that the GSA algorithm produces more accurate partitions than the classical GA using a single software and a single hardware unit. In the same context, in [51], the authors propose a greedy simulated annealing (GSA) algorithm combining the greedy and the SA algorithms. According to the authors this technique improves the performance of the implemented embedded system by an average of 34.96 % and 18.85 % comparing to traditional greedy and SA algorithms. Furthermore, in [61], the authors present an algorithm based on clustering to make GA better in bigger-scale embedded system. It overcomes the shortcoming that algorithm's execution time with the rise number of task-node, to achieve good results in system partitioning. In [20], the authors propose an integration between GA and TS techniques to solve partitioning problem applied to the dynamically reconfigurable system.

Optimal partitioning solution can be performed by applying optimization properties of a partitioning algorithm to increase the performance of an existent one. In [9], the annealing procedure of the SA algorithm is applied to accelerate the updating of the Tabu table in the TS algorithm. The task scheduling is then increased in performance by 50%. Virtual hardware resource is set to implement the customized TS algorithm to improve performance by 97.51%. The combination of Breadth-First–Search (BFS) with Depth-First-Search (DFS) is used for hardware/software task scheduling to fit the features of reconfigurable systems to raise the performance by 50% in comparison with traditional TS and SA algorithms. Furthermore, in [67], the authors use the concept of hardware orientation to create the initial colony of PSO to reduce its randomness. PSO is, then, applied using an updated velocity and position using the concept of the crossover and mutation operators. Finally, TS used the PSO solutions as initial input to find an optimal partition. Results show that the efficiency of this proposed algorithm outperforms comparison algorithm by up to 30% in large-scale problem. Table 3 illustrates some improvement techniques of binary partitioning approach.

TABLE III
SUMMARY OF SOME BINARY IMPROVED PARTITIONING TECHNIQUES

Target Architecture	Algorithms	References
Single software and single hardware components.	Localized Simulated Annealing (LSA) algorithm	[43]
	genetic particle swarm optimization (GPSO) algorithm	[65]
	genetic simulated annealing (GSA) algorithm	[66]
	Restart Hill Climbing (RHC) algorithm	[64]
	The annealing procedure of the SA algorithm is applied to accelerate the updating of the Tabu table in the TS algorithm	[9]
SoC architecture	Advanced Non-Dominated Sorting Genetic Algorithm (ANSGA)	[59]
SoPC	an algorithm to determine the critical path with the largest number of hardware tasks in a given data flow graph	[68]
	Algorithm based on the principle of Binary Search Trees (BST) algorithm and GA	[36]
Reconfigurable System-Architectures	Greedy Simulated Annealing (GSA) algorithm combining the greedy and the SA algorithms	[51]
	Modified GA algorithm by an efficient crossover operator, called DSO	[60]
	Re-excited PSO algorithm	[61]
VLSI two-way architectures	discrete particle swarm optimization (DPSO) with local search strategy, called MDPSO-LS	[62]

In [68], the authors propose a new partitioning algorithm to determine the critical path with the largest number of hardware tasks in a given data flow graph to minimize the area of a SoPC circuit. This technique minimizes the SoPC area (minimize the number of tasks used by the hardware and increase the number of tasks used by the software) while satisfying a time constraint more than SA and GA algorithms.

With the increasing demand of sophisticated functionalities of embedded applications, it is becoming unreasonable to implement them upon uniprocessor architectures. Therefore, embedded systems are frequently implemented today upon multiprocessors architectures. Generally, the embedded systems designer preselects the target architecture in early stage of the design process to reduce the design space. For the systems that consist of multi-processing hardware components and software processing components, partitioning step is very difficult. Such partitioning problem is called extended partitioning.

IV. THE EXTENDED PARTITIONING APPROACH

Traditional partitioning techniques make a binary choice between hardware and software mapping for each application's task (node). Extended partitioning approach can be defined as the cooperation between the determination of mapping (hardware or software), the implementation alternatives (called implementation bins), as well as the scheduling. The techniques developed for binary partitioning approach cannot directly used to tackle the extended partitioning problem with high quality.

Recent studies on extended partitioning approach prove that the partitioning process has a close relationship with scheduling. Generally, researchers on hardware/software co-design for multiprocessors architectures combine partitioning with scheduling [69]. According to [70], the purpose of scheduling is to discover the least-time-cost communication and execution sequence of tasks according to the partitioning from the varied exacts or heuristics algorithms. Tasks scheduling and partitioning on multiprocessors present a NP-hard problems. Different methods are proposed as Scheduling First Partitioning Later (SFPL) [35, 71] and Partitioning First Scheduling Later (PFSL) [6]. For SFPL method, different studies exist. In [71], the authors use an extension of Dijkstra's A-star algorithm for scheduling dependent tasks onto homogeneous processors, which attain better performance using heuristics with cost function. In [35], the authors improve A-star algorithm for scheduling process and introduces benefit-to-area ratio as the priority in partitioning. However, PFSL method are used in the paper [6], in which the authors introduce a new benefit function for partitioning then use Critical-Path and Communication Scheduler (CPCS) algorithm for scheduling. SFPL method compute deeply critical or longest paths, while PFSL distribute hardware/software nodes inserted with greedy scheduling.

Partitioning techniques used in the binary partitioning approach cannot be applied to hardware/software extended partitioning problem. Embedded systems designers have to develop or adjust binary existing techniques to perform an extended partitioning process.

A. New Extended Partitioning Techniques

Several techniques have been presented such as the proposal of [52]. The authors propose a novel multi-level partitioning (MLP) technique to perform hardware-software partitioning in distributed embedded multiprocessor systems (DEMs). This partitioning process consists of three levels. The first level presents a simple binary search allowing quick evaluations of possible partitions. The second level iterates from different possible allocation of software processors to subsystems. The third level iterates over the processors and hardware cost range. Furthermore, In [52], the authors propose a recursive spectral bisection (RSB) for hardware/software partitioning algorithm with time, area and power constraints. Experimental results show that the proposed algorithm is effective for embedded multiprocessor systems. Also, in [53], Recursive spectral bisection (RSB) partitioning technique is proposed to partition hardware and software components to their specified blocks with low communication cost. Authors focused to reduce the execution time, the area resources utilization and the power consumption constraints to target an embedded multiprocessor system. Youness et al. [71] suggest also a new algorithm to reduce the number of processors in homogeneous MPSoC systems combined with hardware FPGA component and reduce the overall execution time and the time-to-market. The used partitioning algorithm depends on the fast conversion of homogeneous software processors that has the longest schedule length to hardware component. This new proposed algorithm was compared with several constructive heuristic techniques to confirm its performance. Recently, Sha et al. [72] propose two algorithms for hardware/software partitioning problem on MPSoC system, to reduce power consumption with time and area constraints: the Tree_Partitioning algorithm which generates optimal solution for tree flow graphs using dynamic programming. The DAG_Partitioning algorithm produces near optimal solution especially for directed-acyclic graphs.

In the same context, Das et al. [73] suggest an optimization technique to choose the partitioning for the software tasks (executed on one or more of the GPPs) and the hardware tasks (implemented on reconfigurable FPGA) to satisfy design cost and performance. Experimental results using synthetic and real application task graphs demonstrate that this technique improve the platform lifetime by 60% as compared to the existing transient fault-aware techniques. Finally, Han et al. [35] present a heuristic algorithm for scheduling and partitioning on MPSoC in order to minimize overall execution time. The proposed algorithm focuses for the critical task graph, and assigns the task with the highest consumption to hardware implementation. Simulation results demonstrate that, the proposed algorithm can decrease execution time up to 38%.

B. Improvement of existing Extended Partitioning Techniques

Other researches choose to improve or adjust existing hardware/software partitioning techniques. TS algorithm are used to partition multiprocessor embedded systems such as in [74]. Authors present a TS on a chaotic neural network, which is a new technique for the low power hardware/software partitioning of heterogeneous target architecture process. They found that the proposed algorithm gets partitioning result with lower energy consumption compared to the GA. The mapped target architecture in our experiments consists of one general processor and two ASICs, and these processing elements are connected through a bus, forming a heterogeneous distributed embedded system.

The hardware-oriented technique is proposed also in [54]. Authors are based on the execution time, memory and area consumption constraints to solve the partitioning problem for embedded multiprocessor FPGA systems using hardware-oriented partitioning technique. They prove the feasibility of their technique by implementing a JPEG encoding system on Xilinx ML310 FPGA platform. Moreover, improvement partitioning in [55] was confirmed by incorporates formal partition with including fitting-system constraints and hardware-oriented partition algorithm. Formal partition is used to rapidly obtain a set of partitioning results that satisfy the system constraints on the number of processor. In [56] paper, the authors propose an efficient hardware/software partitioning for embedded multiprocessor FPGA systems called GHO. This algorithm profit from the advantage of the GA and the hardware-oriented partition for solving partitioning problem of FPGA embedded multiprocessor architecture. This algorithm is also used in [75] partitioning result with faster execution time, smaller memory size and higher slice usage under satisfied system constraints. In the same context, in [76], the authors suggest an evolutionary negative selection algorithm (ENSA-HPS) based on both negative selection model and evolutionary mechanism of the biological Immune System. Results prove that suggested algorithm is more efficient than traditional evolutionary algorithm. In [69], the authors propose an efficient algorithm for dependent task namely Greedy Partitioning and Insert Scheduling Method (GPISM) by task graph. Experimental results demonstrate that GPISM can greatly improve embedded system performance even in the case of generation large communication cost and simplify the partition and the schedule tasks for embedded applications on MPSoC hardware architectures. In [77], the authors suggest a multiple-choice hardware/software knapsack problem (MCKP) based on a TS algorithm and a dynamic programming algorithm. The proposed algorithm based on TS can be applied to solve large partitioning problem o rapidly generate approximate solution, while the dynamic programming algorithm can offer performed solution for small problems. Furthermore, in [78], the authors propose an algorithm based on ACO that simultaneously executes the scheduling, the mapping and the linear placing of tasks, for modern heterogeneous embedded platforms composed of several digital signal, application specific, general purpose processors and reconfigurable devices supporting partial dynamic reconfiguration. Recently, in [79], the authors propose a real-coded genetic algorithm (RCGA) to solve the optimal activation order and the number of processors comparing to simple real-coded GA and PSO algorithms. The proposed algorithm employs a modified crossover and mutation operators to generate a valid solution. The authors employ different population initialization schemes to improve the convergence of the proposed algorithm. In [20], the authors propose a quantum genetic algorithm (QGA) by taking the advantage of quantum computing combined with the traditional GA to reduce the power consumption of the system and decrease the complexity. The scheduling method based on critical tasks was used to solve scheduling problems of task after division. Results verify that the QGA technique increases the diversity of population in the process of task partitioning and the task scheduling algorithm based on the key tasks determines the optimal execution order. The Table 4 summarizes extended partitioning techniques.

TABLE VI
SUMMARY OF SOME EXTENDED PARTITIONING TECHNIQUES

Target Architecture	Algorithms	References
Multiprocessor System Architecture	PFSL: Partitioning First, Scheduling Last Partitioning: Recursive spectral bisection (RSB) has been used to partition. Scheduling: Exchanging tasks between software and hardware components.	[52] [53]
	Partitioning: TS on Chaotic neural network technique	[74]
	PFSL: Partitioning First, Scheduling Last Partitioning: an evolutionary negative selection algorithm (ENSA-HPS) based on both negative selection model and evolutionary mechanism of the biological Immune System.	[76]
	Partitioning: GA combined with Hardware-oriented technique (GHO).	[56]
	Partitioning: real-coded genetic algorithm (RCGA)	[79]
	Partitioning: a quantum genetic algorithm (QGA)	[20]
Heterogeneous MPSoC architecture	PFSL: Partitioning First, Scheduling Last Propose an efficient algorithm for partitioning and scheduling: Greedy partitioning and Insert Scheduling Method (GPISM)	[69]
	Partitioning: The Tree_Partitioning algorithm for tree-structured control-flow graphs using dynamic programming. The DAG_Partitioning algorithm for directed-acyclic graphs.	[72]
	SFPL: Scheduling First, Partitioning last Partitioning: propose technique to reduce execution time by moving the task with the highest benefit-to-area ratio in the critical path iteratively. Scheduling: use A-start algorithm for scheduling	[35]
Homogeneous MPSoC and FPGA	SFPL: Scheduling First, Partitioning last Partitioning: Geometric shape to paths and levels. Path length are calculated and sorted in descending order. Scheduling: A-start as best-first state space search algorithm for scheduling.	[71]
Reconfigurable Architecture: multiple-choice hardware implementation	Partitioning: Efficient heuristic algorithm based on multiple-choice knapsack problem (MCKP) is proposed that is refined by TS algorithm. Dynamic programming algorithm is proposed for the exact solution of the relatively small problems.	[77]
Software processors and reconfigurable devices supporting partial dynamic reconfiguration	an algorithm based on ACO that simultaneously executes the scheduling, the mapping and the linear placing of tasks	[78]

V. Conclusion

The investigation in the Co-design field highlights many motivating areas of studies, especially the hardware/software partitioning process. Hardware/software partitioning presents an enormous challenge for embedded system designers. It consists on dividing an embedded application into software or hardware components generally depending from the application requirements. Several pertinent studies and contributions about hardware/software partitioning techniques and approaches exist. In this paper, we have accomplished a comprehensive review on hardware/software partitioning process from initial specification, the used techniques, to the target implementation architecture. Partitioning process can be classified based on the target architecture as binary and extended approaches. These different techniques are, also reviewed in this paper.

References

[1] G. Sapienza, et al., "Towards a methodology for hardware and software design separation in embedded systems," in the Seventh International Conference on Software Engineering Advances, 2012.
[2] A. Shaout, et al., "Specification and Modeling of HW/SW CO–Design for Heterogeneous Embedded Systems," in Proceedings of the World Congress on Engineering (WCE 2009), 2009.
[3] G. DeMicheli and M. Sami, Hardware/software Co-design vol. 310: Springer Science & Business Media, 2013.
[4] P. L. Marrec, et al., "Hardware, software and mechanical cosimulation for automotive applications," in Rapid System Prototyping, 1998. Proceedings. 1998 Ninth International Workshop on, 1998, pp. 202-206.
[5] A. Bhattacharya, et al., "Hardware software partitioning problem in embedded system design using particle swarm optimization algorithm," in Complex, Intelligent and Software Intensive Systems, 2008. CISIS 2008. International Conference on, 2008, pp. 171-176.
[6] W. Jigang, et al., "Algorithmic aspects for functional partitioning and scheduling in hardware/software co-design," Design Automation for Embedded Systems, vol. 12, pp. 345-375, 2008.
[7] J.-G. Wu, et al., "New model and algorithm for hardware/software partitioning," Journal of Computer Science and Technology, vol. 23, pp. 644-651, 2008.
[8] J. Teich, "Hardware/software codesign: The past, the present, and predicting the future," Proceedings of the IEEE, vol. 100, pp. 1411-1430, 2012.
[9] P. Liu, et al., "Hybrid algorithms for hardware/software partitioning and scheduling on reconfigurable devices," Mathematical and Computer Modelling, vol. 58, pp. 409-420, 2013.
[10] I. Mhadhbi, et al., "High level synthesis design methodologies for rapid prototyping digital control systems," in Programmable Devices and Embedded Systems, 2012, pp. 238-243.
[11] M. López-Vallejo and J. C. López, "On the hardware-software partitioning problem: System modeling and partitioning techniques," ACM Transactions on Design Automation of Electronic Systems (TODAES), vol. 8, pp. 269-297, 2003.
[12] B. Mei, et al., "A hardware-software partitioning and scheduling algorithm for dynamically reconfigurable embedded systems," in Proceedings of ProRISC, 2000, pp. 405-411.
[13] F. Vahid, "Partitioning sequential programs for CAD using a three-step approach," ACM Transactions on Design Automation of Electronic Systems (TODAES), vol. 7, pp. 413-429, 2002.
[14] G. Liang, et al., "A hardware in the loop design methodology for FPGA system and its application to complex functions," in VLSI Design, Automation, and Test (VLSI-DAT), 2012 International Symposium on, 2012, pp. 1-4.
[15] A. Agarwal, "Comparaison of high level design methodologies for arithmetic IPs: Bluespec and C-based synthesis," Master of Science in Electrical Engineering and Computer Science. Massachusetts Institute of technology, 2009.
[16] R. Dömer, et al., "The SpecC System-Level Design Language and Methodology, Part 2," in Parts1 & 2', Embedded Systems Conference, 2002.
[17] J. I. Hidalgo and J. Lanchares, "Functional partitioning for hardware-software codesign using genetic algorithms," in EUROMICRO 97. New Frontiers of Information Technology., Proceedings of the 23rd EUROMICRO Conference, 1997, pp. 631-638.
[18] J. Wu, et al., "Efficient heuristic and tabu search for hardware/software partitioning," The Journal of Supercomputing, vol. 66, pp. 118-134, 2013.
[19] P. Arató, et al., "Algorithmic aspects of hardware/software partitioning," ACM Transactions on Design Automation of Electronic Systems (TODAES), vol. 10, pp. 136-156, 2005.
[20] L. Li, et al., "Hardware/Software Partitioning Based on Hybrid Genetic and Tabu Search in the Dynamically Reconfigurable System," International Journal of Control and Automation, vol. 8, pp. 29-36, 2015.
[21] E. Sha, et al., "Power efficiency for hardware/software partitioning with time and area constraints on mpsoc," International Journal of Parallel Programming, vol. 43, pp. 381-402, 2013.
[22] P. K. Nath and D. Datta, "Multi-objective hardware–software partitioning of embedded systems: A case study of JPEG encoder," Applied Soft Computing, vol. 15, pp. 30-41, 2014.
[23] K. S. Khouri, et al., "Fast high-level power estimation for control-flow intensive design," in Proceedings of the 1998 international symposium on Low power electronics and design, 1998, pp. 299-304.
[24] G. Lin, "An iterative greedy algorithm for hardware/software partitioning," in Natural Computation (ICNC), 2013 Ninth International Conference on, 2013, pp. 777-781.
[25] I. Ahmad, et al., "Multiprocessor scheduling by simulated evolution," Journal of Software, vol. 5, pp. 1128-1136, 2010.
[26] Y.-K. Kwok and I. Ahmad, "Benchmarking and comparison of the task graph scheduling algorithms," Journal of Parallel and Distributed Computing, vol. 59, pp. 381-422, 1999.

[27] H. Topcuoglu, et al., "Performance-effective and low-complexity task scheduling for heterogeneous computing," Parallel and Distributed Systems, IEEE Transactions on, vol. 13, pp. 260-274, 2002.
[28] T. Wiangtong, et al., "Comparing three heuristic search methods for functional partitioning in hardware–software codesign," Design Automation for Embedded Systems, vol. 6, pp. 425-449, 2002.
[29] C. Carreras, et al., "A co-design methodology based on formal specification and high-level estimation," in Proceedings of the 4th International Workshop on Hardware/Software Co-Design, 1996, p. 28.
[30] T.-C. Lin, et al., "Buffer size driven partitioning for HW/SW co-design," in Computer Design: VLSI in Computers and Processors, 1998. ICCD'98. Proceedings. International Conference on, 1998, pp. 596-601.
[31] A. M. Sllame and V. Drabek, "An efficient list-based scheduling algorithm for high-level synthesis," in Digital System Design, 2002. Proceedings. Euromicro Symposium on, 2002, pp. 316-323.
[32] M. Purnaprajna, et al., "Genetic algorithms for hardware–software partitioning and optimal resource allocation," Journal of Systems Architecture, vol. 53, pp. 339-354, 2007.
[33] A. Shaout, et al., "Models of computation for heterogeneous embedded systems," in Electronic Engineering and Computing Technology, ed: Springer, 2010, pp. 201-213.
[34] H. D. Pando, et al., "An application of fuzzy logic for hardware/software partitioning in embedded systems," Computación y Sistemas, vol. 17, pp. 25-39, 2013.
[35] H. Han, et al., "Efficient algorithm for hardware/software partitioning and scheduling on MPSoC," Journal of Computers, vol. 8, pp. 61-68, 2013.
[36] S. Dimassi, et al., "HARDWARE-SOFTWARE PARTITIONING ALGORITHM BASED ON BINARY SEARCH TREES AND GENETIC ALGORITHM TO OPTIMIZE LOGIC AREA FOR SOPC," Journal of Theoretical & Applied Information Technology, vol. 66, 2014.
[37] P. Eles, et al., "Hardware/software partitioning with iterative improvement heuristics," in Proceedings of the 9th international symposium on System synthesis, 1996, p. 71.
[38] J. Axelsson, "Architecture synthesis and partitioning of real-time systems: A comparison of three heuristic search strategies," in Proceedings of the 5th International Workshop on Hardware/Software Co-Design, 1997, p. 161.
[39] K. S. Chatha and R. Vemuri, "Hardware-software partitioning and pipelined scheduling of transformative applications," Very Large Scale Integration (VLSI) Systems, IEEE Transactions on, vol. 10, pp. 193-208, 2002.
[40] I. Mhadhbi, et al., "Impact of Hardware/Software Partitioning and MicroBlaze FPGA Configurations on the Embedded Systems Performances," in Complex System Modelling and Control Through Intelligent Soft Computations, ed: Springer, 2015, pp. 711-744.
[41] M. O'Nils, et al., "Interactive hardware-software partitioning and memory allocation based on data transfer profiling," in Proc. International conference on Recent Advances in Mechatronics, 1995.
[42] J. Wu and T. Srikanthan, "Low-complex dynamic programming algorithm for hardware/software partitioning," Information Processing Letters, vol. 98, pp. 41-46, 2006.
[43] T. Wiangtong, et al., "Comparing Three Heuristic Search Methods for Functional Partitioning in Hardware–Software Codesign," Design Automation for Embedded Systems, vol. 6, pp. 425-449, 2002/07/01 2002.
[44] Y.-K. Kwok, et al., "A semi-static approach to mapping dynamic iterative tasks onto heterogeneous computing systems," Journal of Parallel and Distributed Computing, vol. 66, pp. 77-98, 2006.
[45] F. Ghaffari, et al., "Algorithms for the Partitioning of Applications containing variable duration tasks on reconfigurable architectures," in Computer Systems and Applications, 2003. Book of Abstracts. ACS/IEEE International Conference on, 2003, p. 13.
[46] G. Stitt and F. Vahid, "Energy advantages of microprocessor platforms with on-chip configurable logic," IEEE Design & Test of Computers, pp. 36-43, 2002.
[47] G. Stitt and F. Vahid, "A decompilation approach to partitioning software for microprocessor/FPGA platforms," in Design, Automation and Test in Europe, 2005. Proceedings, 2005, pp. 396-397.
[48] T. Streichert, et al., "Online hardware/software partitioning in networked embedded systems," in Proceedings of the 2005 Asia and South Pacific Design Automation Conference, 2005, pp. 982-985.
[49] P. Waldeck and N. Bergmann, "Dynamic hardware-software partitioning on reconfigurable system-on-chip," in System-on-Chip for Real-Time Applications, 2003. Proceedings. The 3rd IEEE International Workshop on, 2003, pp. 102-105.
[50] P. K. Sahu, et al., "Extending Kernighan–Lin partitioning heuristic for application mapping onto Network-on-Chip," Journal of Systems Architecture, 2014.
[51] Y. Jing, et al., "Application of improved simulated annealing optimization algorithms in hardware/software partitioning of the reconfigurable system-on-chip," in Parallel Computational Fluid Dynamics, 2014, pp. 532-540.
[52] L. Trong-Yen, et al., "Hardware-software multi-level partitioning for distributed embedded multiprocessor systems," IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, vol. 84, pp. 614-626, 2001.
[53] T.-Y. Lin, et al., "Efficient hardware/software partitioning approach for embedded multiprocessor systems," in VLSI Design, Automation and Test, 2006 International Symposium on, 2006, pp. 1-4.
[54] T.-Y. Lee, et al., "Hardware-oriented partition for embedded multiprocessor FPGA systems," in Innovative Computing, Information and Control, 2007. ICICIC'07. Second International Conference on, 2007, pp. 65-65.
[55] T.-Y. Lee, et al., "Enhancement of hardware-software partition for embedded multiprocessor fpga systems," in Intelligent Information Hiding and Multimedia Signal Processing, 2007. IIHMSP 2007. Third International Conference on, 2007, pp. 19-22.
[56] T.-Y. Lee, et al., "An Efficiently Hardware-Software Partitioning for Embedded Multiprocessor FPGA Systems," presented at the International Multiconference of Engineers and Computer Scientists (IMECS), Hong Kong, 2007.
[57] T.-Y. Lee, et al., "Sophisticated Computation of Hardware-Software Partition for Embedded Multiprocessor FPGA Systems," in Innovative Computing Information and Control, 2008. ICICIC'08. 3rd International Conference on, 2008, pp. 81-81.
[58] S.-q. Luo, et al., "An Advanced Non-Dominated Sorting Genetic Algorithm Based SOC Hardware/Software Partitioning [J]," Acta electronica sinica, vol. 11, p. 042, 2009.
[59] S. G. Li, et al., Hardware/Software Partitioning Algorithm Based on Genetic Algorithm vol. 9, 2014.
[60] R. Faraji and H. R. Naji, "An efficient crossover architecture for hardware parallel implementation of genetic algorithm," Neurocomputing, vol. 128, pp. 316-327, 2014.
[61] M. B. Abdelhalim and S.-D. Habib, "An integrated high-level hardware/software partitioning methodology," Design Automation for Embedded Systems, vol. 15, pp. 19-50, 2011.

[62] W. Guo, et al., "A hybrid multi-objective PSO algorithm with local search strategy for VLSI partitioning," Frontiers of Computer Science, vol. 8, pp. 203-216, 2014/04/01 2014.
[63] P. Xiao, et al., "Hardware/software partitioning based on improved simulated annealing algorithm," Journal of Computer Applications, vol. 7, p. 019, 2011.
[64] H. D. Pando, et al., "An application of fuzzy logic for hardware/software partitioning in embedded systems," Computación y Sistemas, vol. 17, pp. 25-39, 2013.
[65] L. An, et al., "Algorithm of hardware/Software partitioning based on genetic particle swarm optimization," Journal of Computer-Aided Design & Computer Graphics, vol. 6, p. 005, 2010.
[66] X. Zhao, et al., "An Effective Heuristic-Based Approach for Partitioning," Journal of Applied Mathematics, vol. 2013, 2013.
[67] L. Guoshuai, et al., "Improved Hardware/software Partitioning Algorithm Based on Combination of PSO and TS," Journal of Computational Information Systems, vol. 10, pp. 5975–5985, July 15, 2014 2014.
[68] M. Jemai and B. Ouni, "Hardware software partitioning of control data flow graph on system on programmable chip," Microprocessors and Microsystems, vol. 39, pp. 259-270, 2015.
[69] C. Li, et al., "A dependency aware task partitioning and scheduling algorithm for hardware-software codesign on MPSoCs," in Algorithms and Architectures for Parallel Processing, ed: Springer, 2012, pp. 332-346.
[70] C. Boeres and V. E. Rebello, "Towards optimal static task scheduling for realistic machine models: Theory and practice," International Journal of High Performance Computing Applications, vol. 17, pp. 173-189, 2003.
[71] H. Youness, et al., "A high performance algorithm for scheduling and hardware-software partitioning on MPSoCs," in Design & Technology of Integrated Systems in Nanoscal Era, 2009. DTIS'09. 4th International Conference on, 2009, pp. 71-76.
[72] E. Sha, et al., "Power efficiency for hardware/software partitioning with time and area constraints on mpsoc," International Journal of Parallel Programming, pp. 1-22, 2013.
[73] A. Das, et al., "Aging-aware hardware-software task partitioning for reliable reconfigurable multiprocessor systems," in Proceedings of the 2013 International Conference on Compilers, Architectures and Synthesis for Embedded Systems, 2013, p. 1.
[74] T. Ma, et al., "Low power hardware-software partitioning algorithm for heterogeneous distributed embedded systems," in Embedded and Ubiquitous Computing, ed: Springer, 2006, pp. 702-711.
[75] T.-Y. Lee, et al., "Hardware-software partitioning for embedded multiprocessor FPGA systems," International Journal of Innovative Computing, Information and Control, vol. 5, pp. 3071-3083, 2009.
[76] Y. Zhang, et al., "A hardware/software partitioning algorithm based on artificial immune principles," Applied Soft Computing, vol. 8, pp. 383-391, 2008.
[77] J. Wu, et al., "Algorithmic aspects for multiple-choice hardware/software partitioning," Computers & Operations Research, vol. 39, pp. 3281-3292, 2012.
[78] F. Ferrandi, et al., "Ant Colony Optimization for mapping, scheduling and placing in reconfigurable systems," in Adaptive Hardware and Systems (AHS), 2013 NASA/ESA Conference on, 2013, pp. 47-54.
[79] S. Suresh, et al., "Hybrid real-coded genetic algorithm for data partitioning in multi-round load distribution and scheduling in heterogeneous systems," Applied Soft Computing, vol. 24, pp. 500-510, 2014.

Heterogeneous Embedded Network Evaluation of CAN-Switched ETHERNET Architecture

Nejla REJEB[1], Ahmed Karim BEN SALEM[2], Slim BEN SAOUD[3]
LSA Laboratory, INSAT-EPT, University of Carthage, TUNISIA

Abstract-The modern communication architecture of new generation transportation systems is described as heterogeneous. This new architecture is composed by a high rate Switched ETHERNET backbone and low rate data peripheral buses coupled with switches and gateways. Indeed, Ethernet is perceived as the future network standard for distributed control applications in many different industries: automotive, avionics and industrial automation. It offers higher performance and flexibility over usual control bus systems such as CAN and Flexray. The bridging strategy implemented at the interconnection devices (gateways) presents a key issue in such architecture. The aim of this work consists on the analysis of the previous mixed architecture. This paper presents a simulation of CAN-Switched Ethernet network based on OMNET++. To simulate this network, we have also developed a CAN-Switched Ethernet Gateway simulation model. To analyze the performance of our model we have measured the communication latencies per device and we have focused on the timing impact introduced by various CAN-Ethernet multiplexing strategies at the gateways. The results herein prove that regulating the gateways CAN remote traffic has an impact on the end to end delays of CAN flow. Additionally, we demonstrate that the transmission of CAN data over an Ethernet backbone depends heavily on the way this data is multiplexed into Ethernet frames.

Keywords: Ethernet, CAN, Heterogeneous Embedded networks, Gateway, Simulation, End to end delay.

I. INTRODUCTION

Embedded network architectures on transportation systems are currently witnessing major changes. Aircraft and vehicle tend to be more electronic with a larger use of on-board microprocessors. Interconnection equipments for heterogeneous networks become popular in automotive and recently in avionics context. These devices allow various communication standards and equipments to coexist with each other. Data flow between systems and the number of connections between functions will, therefore, increase. Thus, new necessities in embedded networks have appeared.

The use of Ethernet is currently investigated in many industries [1, 2], such as automotive, avionics and industrial automation. Two important advantages have been noticed from using Ethernet. Firstly, it is flexible and endowed with an open standard with no tight bounds to a specific supplier. Secondly, it offers high data rates at an extremely low price point compared to domain-specific protocols like Controller Area Network (CAN) [3], FlexRay [4] etc. However, the predictable timing of data transfers is perceived as a major challenge when using Ethernet in industrial sectors. Furthermore, currently networks are complex heterogeneous systems. They consist of different sub-heterogeneous networks (field busses) interconnected to the federator technology Switched ETHERNET. This new technology (multiplexed communication networks based on Ethernet) presents new challenges towards the time criticality of information circulating. It is therefore necessary to design and develop solutions to this type of network to satisfy the Quality Of Service QoS constraints and requirements (length of stay of a frame, latency, etc.). For distributed control applications, predictable communication delay is highly important and can be problematic in case of using standard Ethernet. Therefore, in order to handle heterogeneity between ETHERNET backbone and peripheral data buses, modular gateways [5] dispersed all over the aircraft or vehicle, are used. Gateways, actually, become the main node that will need reconfiguration. They are among major challenges in the design process of such multi-cluster networks. They include the performance analysis of the bridging strategy between the different technologies. Thus, new study techniques are required for design certification and network performance analysis. Hence, this study focuses on real-time performance evaluation of heterogeneous embedded networks. We consider a heterogeneous network architecture that is already integrated into the aircraft and vehicle: Switched ETHERNET-CAN. Messages are transmitted by more than one technology. Thus, we analyze, in this paper, the end-to-end delays over such a heterogeneous path. . A gateway has to achieve fast data exchange between the ETHERNET network and CAN busses. So the signals packed in received messages on each network have to be mapped to be transmitted on the other network with a bounded processing delay. This article focuses on the timing impact introduced by various CAN/Ethernet multiplexing strategies at the gateways. It is organized as following: section 2 and 3 present an overview of related works and heterogeneous networks. Section 4 analyze the case study of CAN-Ethernet and propose its gateway model. In the section 5, we study and compare different bridging strategies based on OMNET++ simulation tools [6]. Finally, the section 6 puts forward the conclusion of this study and presents some ideas for future works.

II. HETEROGENEOUS NETWORKS

The new networks architecture include different sub-heterogeneous networks (field busses, traditional avionics protocols such as ARINC 429 [7], traditional automotive embedded networks such as CAN [3] and Flexray[4], sensor networks, open world network, etc.), that are related to the Switched ETHERNET federator technology. To support the deterministic industrial communications by ensuring bounded end-to-end delays, several solutions have been presented. In the context of avionics, Avionics Full DupleX (AFDX) switched Ethernet, was introduced. Recently new aircrafts have a whole new architecture that incorporates different fields, applications, and heterogeneous networks. Typical avionics network architecture is shown in Fig.1.

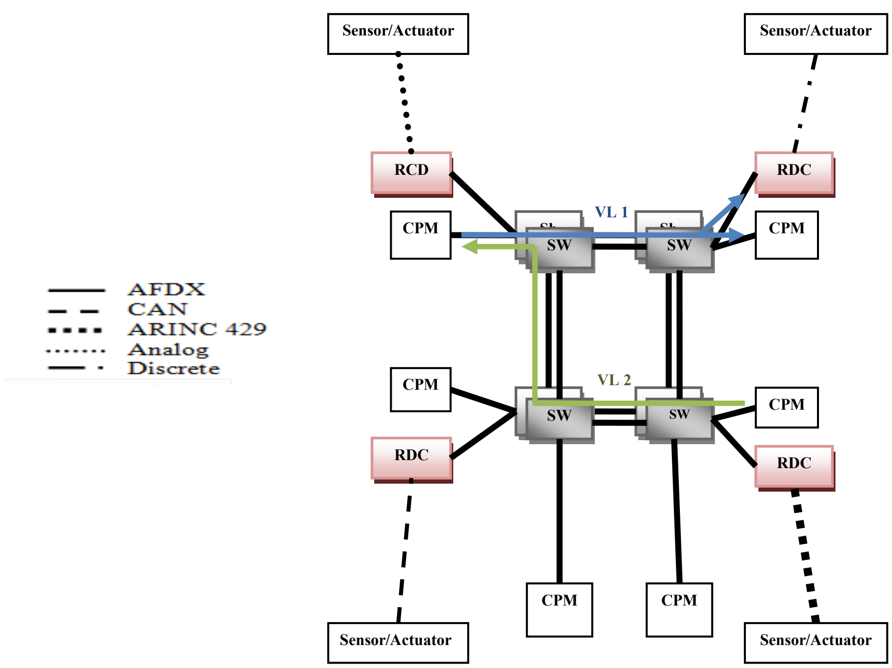


Figure 1. Avionic heterogeneous embedded network [7, 8]

This Avionics Data Communication Network (ADCN) is composed of: (i) Shared processing units: modules in charge of the execution of applications (Core Processing Modules (CPM), Line Replaceable Units (LRU)), (ii) Multiplexed avionics communications network : deterministic switched Ethernet connected by AFDX SWitches (SW), (iii) Elements located outside the Integrated modular avionics (IMA) [9], connected to the avionic world by field busses (analog, discrete, ARINC 429, CAN), (iv) GateWays (GW) modules (Remote Data Concentrator (RDC [5])) for messages transmission between the AFDX basic network and the peripheral communication busses. These devices are necessary to handle dissimilarities of protocol and guarantee interoperability between AFDX and other field busses. New requirements in embedded heterogeneous networks have emerged. Interconnection devices should be designed to ensure correct protocol conversion between network clusters, guaranty real-time requirements, and improve network efficiency. The literature encloses many pertinent studies about embedded heterogeneous networks. Both performance evaluation of these heterogeneous networks and interconnection device design present the main subject of these studies.

III. Related works

New avionics and automotive systems are tightly related to multiplexed communication networks such as Ethernet. Mechanisms ensure a minimum bandwidth for each data stream, and can limit the transit time of each message crossing the network. Performance evaluation of these networks will be necessary. In the area of optimizing resources over critical embedded networks, different approaches have been proposed and included into the end-to-end communication delay. This includes traffic-sources, interconnection devices and communication networks for embedded networks in order to guarantee higher timing performances and allow a better load balance in the network. We have been interested in two most used metrics in the literature and we are exposing them in this related works: timing analysis technique for embedded networks and bridging strategies on heterogeneous networks.

A. Timing verification approaches

For certification reasons, it is necessary to prove that the communication delay for each message does not exceed its deadline, in automotive and avionics networks. To accomplish this, in the literature, various methods have been proposed. As example of methods to compute end-to-end communication latencies and analyze the timing performances of critical embedded networks, a network calculus [10-12], trajectory approach [13], model checking [14] and simulation [15] have been introduced. Two essential groups of approaches can be distinguished: the first group is based on analytic theories examination and the second one is based on simulation approach. For any specified network, they provide solutions to compute an upper bound on end-to-end communication delays or to compute exact end-to-end communication. Analytic methods are based on mathematical models. **The Network Calculus** approach [16, 17] which is based on the MinPlus algebra theory [18] to compute upper bounds, has been commonly used for analyzing performances in computer networks. For avionic context, it was one of the first methods used for AFDX certification. It gives safe but pessimistic upper-bounds on the end-to-end delays of flows. This approach can indeed lead to impossible scenarios. More recently, **the trajectory approach** has been proposed as a response to time analysis [15]. This approach has been applied to the AFDX context in [19]. It is a timing analysis technique introduced to get deterministic upper bounds on communication response times in distributed systems .It is based on the analysis of the worst-case scenario experienced by a packet on its trajectory. It also gives safe upper-bounds which are slightly tighter compared with the network calculus approach. Another approach to study the maximum end-to-end delay is **the Model checking** [20] . This method determinates the exact worst-case transmission delays. This Model Checking approach checks all the possible scenarios that can be experienced on the network to find the exact worst-case end-to-end delays. A performance analyze of AFDX network was done in [24] by applying the Model Checking approach. In the general case, finding this worst-case scenario requires an exhaustive analysis of all the possible scenarios. Such an exhaustive enumeration is limited by the combinatorial explosion problem, since the number of possible scenarios is huge. Whereas, the goal of **the simulation approach** presented in [21, 22] is to approximate real network behavior. It is a very interested approach that involves a network for designing and evaluating. It needs a realistic model of network and calculates the end-to-end delay of a given flow on a subset of all possible scenarios. The guided simulation approach seeks to assess the pessimism bounds calculated using network calculus in determining a distribution of end to end delay. But, to certify the communication determinism required by critical embedded networks, the simulation approach is not sufficient. So, **the stochastic Network Calculus approach** [19] can be used as a complementary method to the simulation and the calculation of deterministic networks. The resulting distribution is pessimistic compared with the actual behavior of the network calculated by the model checking and estimated by a simulation approach, but much less pessimistic than the upper bound obtained by the deterministic network calculus approach.

For our work, analytical or simulation methods can be used to valuate CAN–ETHERNET network. But, simulation methods have been retained for the sake of this study since we notice that they are effective for a better understanding of heterogeneous network behavior and they allow a better real-time performance evaluation (end to end delay, jitter, etc.).

B. Bridging strategies

Bridging equipments (GW) for heterogeneous networks have become commonly used in automotive and recently in automotive and avionics context. These GWs permit the coexistence of different communication networks and equipments. The examination of timing impact of these GWs and their bridging strategies will require a multitude of analysis methods. In automotive and avionics context, relating to the interconnection equipment design, literature [23-26] study several timing performance evaluation approaches and propose different bridging strategies. The three strategies have been compared by simulation in [24]. The results show that concerning end to end delays, the timed n for one strategy gives the best ratio of CAN frames. In [28], authors introduced a new heterogeneous CAN-Ethernet architecture to interconnected CAN buses. Ayed et al. Proposed, in [27], a similar strategy regarding AFDX specific characteristics. Communication happens across Virtual Links (VLs), which has a reserved bandwidth and a maximum frame size. These strategies are classified into three types. **The one for one strategy:** A more straightforward encapsulation strategy that puts each global CAN frame in a separate Ethernet frame and transmits it shortly. Here, we notice that there are time limits for CAN frames once non-CAN Ethernet load is higher than or equal to 20 Mbs. **The n for one strategy:** This strategy consists in encapsulating *n* CAN frames in an Ethernet frame (frame bunching), which implies that each global CAN frames has to halt until pending CAN frames in the bridge station come along. **The timed n for one strategy**: In order to improve results, we have to guarantee that no global CAN frame will defer more than a specific amount of time before being encapsulated and send to Ethernet.

IV. CAN-SWITCHED ETHERNET ARCHITECTURE

The retained heterogeneous communication architecture is represented by Fig.2, that consists of the following sub-systems: (i) Full Duplex Switched network (Ethernet End Systems interconnected by Ethernet Switch); (ii)GW that allows communication between the Ethernet world and the CAN peripheral network (sensor network, open world, etc.); and (iii) CAN busses (used for data exchange from sensors or to actuators). Nowadays, in addition to the automobile environment, CAN bus is employed as one of the main avionics networks in general aviation architectures. It is useful for linking sensors, actuators and other types of avionics devices that are usually necessary for low medium data transmission. Adding to that, Switched Ethernet and CAN bus are among the most promising technologies available for the aerospace industry and automotive domain. Indeed, they provide a large bandwidth and a network structure that will allow wiring reduction and guarantee high reliability at the same time. For the following reasons, we have chosen this architecture.

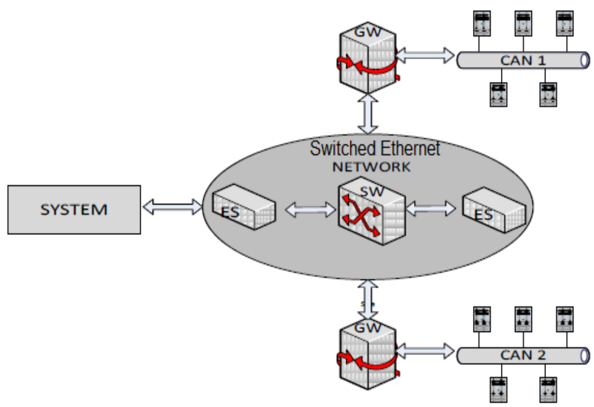


Figure 2. Heterogeneous Switched Ethernet-CAN

A. *Communication technologies*

1) *Full Duplex Switched Ethernet*

Carrier Sense Multiple Access/ Collision Detection (CSMA/CD) is the Ethernet native medium access method. The collision resolution mechanism is described as non deterministic and leads to unbounded transmission delays. Full Duplex Switched Ethernet [28] is a way to bypass the drawbacks of this mechanism : The physical link collisions are eliminated since each station is tightly related to an Ethernet switch with a complete full duplex link. Consequently, guaranteed performances are strongly related to switch policies. Therefore, such Full Duplex Switched Ethernet network receives an increasing attention in the industrial domain. In this paper, we consider a very basic switch with a First-In First-Out policy in each output port. The Ethernet frame format for 10Mbs and 100Mbs is presented in Table I. Ethernet frame is composed of (at least) 26 bytes of control and 0 up to 1500 bytes of data.

TABLE I. ETHERNET FRAME FORMAT (LENGTH IN BYTE)

7	1	6	6	2	0-1500	4
Preamble	SOF	Destination Address	Source Address	Type	Data + JUMP	FCS

Additionally, we note that AFDX network is an example of real-time switched Ethernet network. It is defined in the context of avionics and developed for modern aircraft such as Airbus A380 (IEEE 802.3 and ARINC 664, Part 7). AFDX technology for avionics [29] improves higher data speed transfer and reduces wiring. As a result, determinism is enhanced and bandwidth is guaranteed.

2) Controller Area Network (CAN)

CAN bus has been successfully used for decades in automotive due to its high reliability, real-time properties and low cost. It becomes recently an attractive communication technology for aircraft manufacturers. Recent CAN standards have been introduced for avionics, such as CAN Aerospace and ARINC 825 [30]. CAN [31] is an asynchronous multi-master serial data bus that was standardized in 1993 [32]. CAN is able to operate at speeds of up to 1 Mbit/s with a payload message of at most 8 bytes and an overhead of 6 bytes due to the different headers and bit stuffing mechanism. The data transmitted on CAN bus are packed into unique CAN identifiers message frames (CAN IDs). A Carrier Sense Multiple Access / Collision Resolution (CSMA/CR) protocol is used to resolve the collisions on the bus. When two or more CAN stations start a transmission at the same time, the one with the highest priority (lowest value) wins and the others stop their transmission. This is implemented by collision detection thanks to the bit arbitration method. The CAN frame format [32] is depicted in Table II. The following fields, mentioned in Table II, are important for the study: (i) Identifier field identifies the data contained in the frame. In our work, we have considered a standard CAN frame with 11-bit ID. (ii) DLC gives the data length. (iii) Data field consists of the frames payload.

TABLE II. CAN FRAME FORMAT (LENGTH IN BITS) [25]

1	11/29	1	1	1	4	0-64	16	2	7	3
SOF	Identifier	rtr	IDE	r0	DLC	Data	CRC	ACK	EOF	IFS

3) CAN-ETHERNET gateway

This device is necessary to handle the dissimilarities between CAN and Ethernet in terms of communications and protocol characteristics: The available bandwidth (1MBs or less for CAN, 100 MBs for Ethernet), the addressing system (ID associated to data for CAN, MAC Addresses of station for Switched Ethernet), different Maximum Transfer Unit (MTU) (data size between 0 and 8 bytes for CAN and between 0 and 1500 bytes for Ethernet) and the collision resolution (CSMA/CD for Ethernet and CSMA/CR for CAN).

B. Network topology

Our adopted network is a CAN-Ethernet network representing an existing avionic or automotive system. Several CAN busses are interconnected with a full duplex switched Ethernet network in order to evaluate the end to end delay of the circling information in this heterogeneous network. We have considered the same network architecture studied in [24] and depicted in Fig.3 in order to compare its performance with our own designed architecture. Fig 3 presents an illustrative network topology that includes 3 CAN busses and an Ethernet switch. There is a bridge station between each CAN bus and the switch. The switch has three reception ports and three queued transmission ports. When a frame arrives at the switch, the control logic determines the port and tries to transmit the frame immediately. If the port is busy, the frame is stored in the FIFO out of transmission port queue. The switching tables are static since all flows are previously recognized in this kind of architecture.

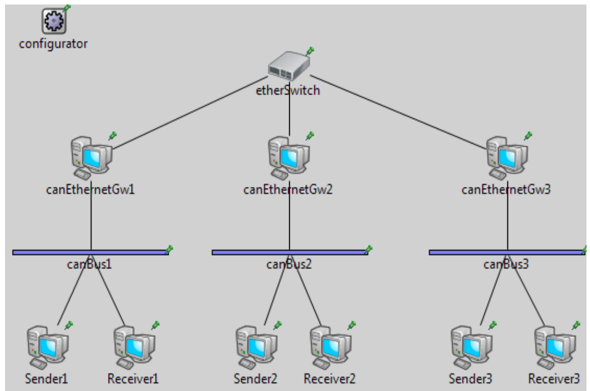


Figure 3. Adopted Network topology

C. *Network traffic description*

We have considered the same traffic described in [24]. 42 messages are transmitted on this architecture as described in Table III. We study the case where a CAN network and an Ethernet network exchange messages via a GW. Both CAN bus 1 and CAN bus 3, generates 15 local frames and 1 remote frame. However, CAN bus 2 has 10 local frames and receives 2 remote frames originated from CAN bus 1 and CAN bus 2. These remote messages have the lowest priority on their source bus and the highest one on their destination bus.

In the following sections, we will show the impact of different bridging strategies on this specific scenario. In order to evaluate our network case study, we have chosen a simulation environment based on the discrete Event Simulator OMNET ++[6].

TABLE III. NETWORK TRAFFIC DESCRIPTION

Identifiers	Pi (ms)	DLC (byte)	Src bus	Dest bus	Flow type
1, 3 ,7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29	4	8	1	1	Local
5	2	2	1	1	Local
38, 39, 40, 41, 42, 43, 44, 45, 46, 47	2	8	2	2	Local
2, 4, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30	2	8	3	3	Local
6	2	2	3	3	Local
31	2	8	1	2	Remote
32	2	8	3	2	Remote

We have considered an Ethernet link at 100Mbs and all the CAN busses correspond to 1 Mbs. For the CAN bus, we have used a model already integrated on the OMNET++. This model was designed and validated by Matsumura and al. in [21] . But to evaluate our network, we have developed our own CAN-Ethernet GW model. We will present the design details of this bridging equipment in the next section.

D. CAN/ETHERNET gateway design

Since global CAN traffic has to be transmitted on the Ethernet network, it is necessary to define a bridging strategy, to handle dissimilarities between Ethernet and CAN bus. GW allows communication between two different architectures and protocols. The choice of an encapsulating policy is judged judicious due to the difference in characteristics between CAN and Ethernet. This GW must perform a conversion protocol between both networks. It has to extract the payloads of the received messages and add the correct protocol headers before transferring them to their destination. Therefore, an appropriate data formatting strategy is required to be implemented on the GW to satisfy the requirements of destination network. In addition, a routing function will be needed in the GW. Therefore, each network includes addressing scheme mapped to physical addressing scheme. CAN have to adopt the mapping function of Ethernet network and vice versa. Therefore, our GW model comprises two functions: a messages router and CAN-UDP protocol converter (changing CAN messages to UDP packets and vice versa). The Fig.4 illustrates the protocol structure of the considered GW.

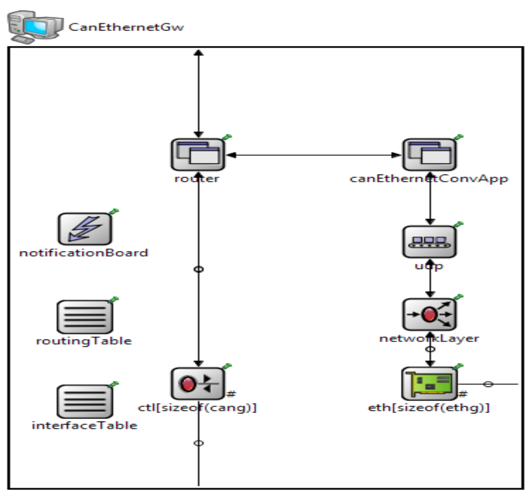


Figure 4. CAN-ETHERNET GW model on Omnet++ [21]

A distinction is made between traffic from CAN to Ethernet and Traffic from Ethernet to CAN as explained by Fig.5. The Identifier and the payload are extracted after the decapsulation of each received CAN frame. Then, this data is encapsulated in the data field of the Ethernet frame. The encapsulation consists in putting the ID (11bits) and Data fields (64 bits) of CAN frames in the Data field (0 to 1500 bytes) of the Ethernet frame. This means that CAN frame will occupy at most 10 bytes of the DATA field on an Ethernet frame [33]. So, many CAN frames can be assembled at the GW node and sent on the same Ethernet MAC address on the Ethernet network towards the ultimate destination. Therefore, the routing table associates one MAC address to many CAN identifiers [33]. However, concerning the traffic from Ethernet to CAN, a fragmentation process is required to regulate and adjust the MTU frame size differences, since generally the payload of an Ethernet message is larger than the maximum payload of CAN messages. This fragmentation process occurs at the GW node and at the CAN destination, a reassembling will take place. Each fragmented frame format will be the same as the original frame format while making the only exception of the network MTU size over which it will transfer. Then, according to the mapping function, each received Ethernet frame that is fragmented into multiple CAN frames is routed. In our case, for each CAN identifier there is an associated Mac address on the Ethernet network [33]. So a static mapping table is considered.

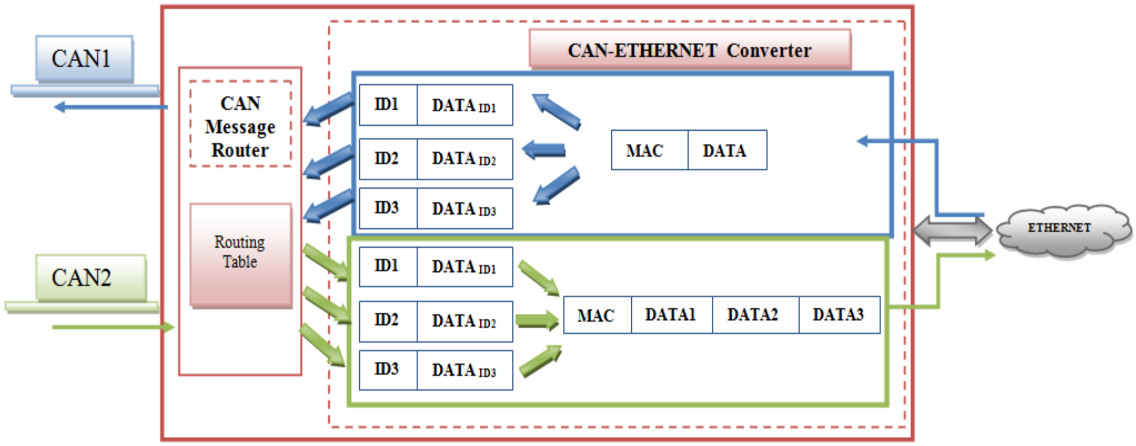


Figure 5. Functionality conversion

V. Gateway Impact Study

The examination and the analysis of the GWs characteristics and their impact on the performance of end-to-end delay are a major challenge in the design process of heterogeneous embedded systems. An important performance metric for a GW is the processing delay, that is defined as the difference between the signal transmit time from the GW and the same signal received time in the GW. This processing delay parameter is a component among others of the end-to-end delay of the signals conveyed in the message.

A. *Gateway model validation*

In order to design our own model, we have relied on the GW model described in [21]. We have validated our model of Fig.4 by testing a simple network composed by a CAN bus and an Ethernet node interconnected by a CAN-ETHERNET bridge. Thanks to this example, we are able to test and estimate the various conversion features. Then, we simulate a heterogeneous network CAN1-GW1-SW-GW2-CAN2 before testing the entire scenario described in table III. Fig.6 presents the simulation results illustrated by the percentage of latencies of each component over such a heterogeneous path.

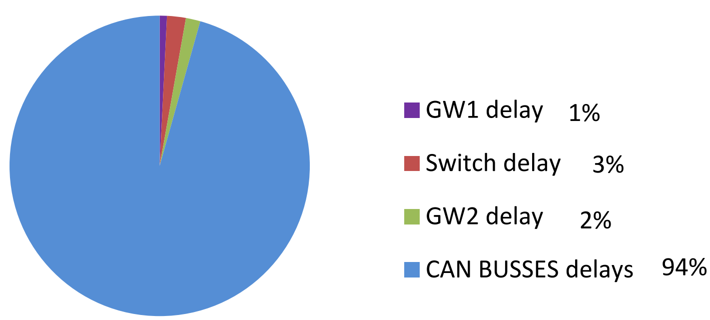


Figure 6. Component's latencies in a heterogeneous path

The global end-to-end delay on a CAN- Switched Ethernet network may be determined as following:

$$Dglobal = \sum_i D_{CANi} + \sum_j D_{GWj} + \sum_k D_{SWk} \quad (1)$$

Where*:*

- D_{CANi} is latency of the ith CAN bus;
- D_{GWj} is latency of the jth GW*;*
- D_{SWk} is latency Ethernet Switches.

The GW delays are equal to the payload extraction and mapping latency. The duration of the message latency at the GW is affected by the GWs mapping strategy according to their functions. So, the determination of such a delay is necessary for the end to end delay evaluation of a global system.

The latency on the GW may be defined as:

$$D_{GW} = D_{Rx} + D_{O.GW} + D_{Tx} \quad (2)$$

Where:

- D_{Rx} is the delay needed for an incoming message until the message is served from the input buffer;
- $D_{O.GW}$ is the GW operating time;
- D_{Tx} is the delay needed until an outgoing message on the output buffer can be sent in the destination domain.

The transmission delay of a frame through the CAN bus depends on a wide range of conditions presents on the bus and the messages currently being exchanged (the number of data bytes transmitted, the presence or absence of error frames, etc.).

We examine the worst possible case, and try to calculate the transmission time of a CAN message. We consider that the network is not subject to disturbance (no error frame).

The worst case arbitration delay is computed as following:

$$D_{CAN} = {}^{L}/_{B} \quad (3)$$

Where:

- L is the length of longest message: is calculated according to the work of Davis et al.[31]. It considers the worst case overhead induced by bit stuffing using:

$$L = 47 + 8 \times DLC + \left(\frac{34 + 8 \times DLC}{4}\right) \quad (4)$$

- B is the Baude rate.

We have used a CAN busses with 1 Mbs. The transmission time is 135 μs for a message of 8 bytes according to (3). We have confirmed this value by simulation.

B. Gateway strategies impact

1) One to one bridging strategy

We will focus on the following in testing different GWs strategies to show their impact on real time behavior of the network. The GW decapsulates the included remote CAN frames in it and transmits it on the CAN bus, when it receives an Ethernet frame. We have proposed two approaches in this paper:

a) Immediate forwarding strategy

In this section, we will consider the one-to-one strategy and we will assume that each CAN frame is encapsulated in a separate Ethernet frame, since there is only one remote frame generated by each bus. We presume that the offsets of all flow are null. The most straightforward encapsulation strategy is to insert each global CAN frame in a separate Ethernet frame and transmit it promptly. The remote frames from CAN1 or CAN3 are encapsulated in an Ethernet frame and sent straight away to the GW of CAN bus 2. Since there are no competing frames at that instant, frame is then decapsulated and ready instantly for transmission on CAN bus 2. Indeed, the immediate forwarding of remote CAN frames by their GW can decrease considerably the time length between the instants where two successive messages of a specific CAN flow get ready on their destination bus. Moreover, such a burst of traffic can also increase the delay of the local flows of CAN bus 2.

We have accomplished the simulation for two scenarios, with and without remote frames. Fig.7 represents the end to end delay for each frame in CAN 2 (local and distant) during two periods for each scenario. We noticed that after arriving at the bus 2 at 2 ms (Cycle 2) remote frames, as they are priority (Ids= 0x31 and 0x32), are transmitted before the local frames. Therefore, they affect the latencies and increase the end to end delays of local frames. For example, the end to end delay of the local frame of ID 0x47 Pass from 3.35 ms to 3.93 ms. Moreover, a comparison between our results and those presented in [27], using the same network architecture and GW strategy, is made in order to prove the efficiency of our model in term of delay. Thus, we confirm the analysis presented in [27] by simulation using OMNET++ simulation tool. The two results show that almost local frames had missed their deadline. This means, that remote frames have an impact on the end to end delays of a local bus.

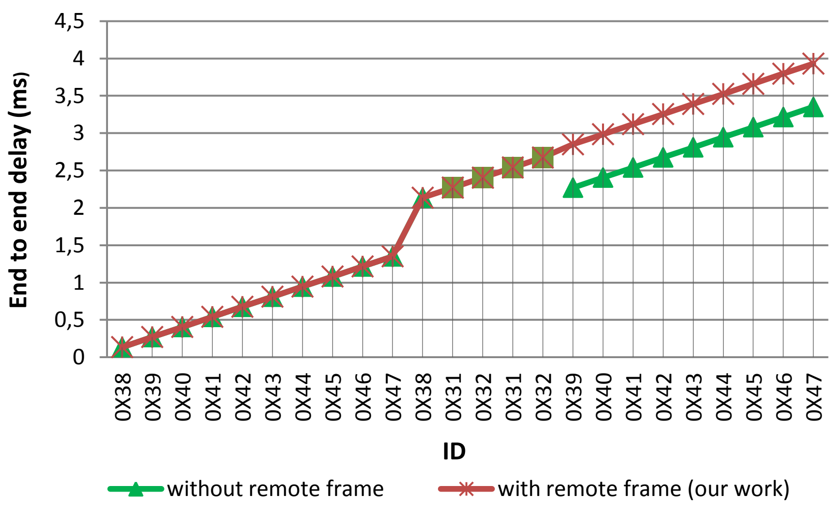


Figure 7. Frames end to end delays (CAN2) (local and remote frames)

Fig.8 illustrates the difference (delay introduced by remote frames) between the two delays measured in Fig.7. This problem can be solved, if we ensure that distant CAN frame will be deferred for a specific amount of time on GW before transmitted to their destination. A second strategy has been proposed in the next section.

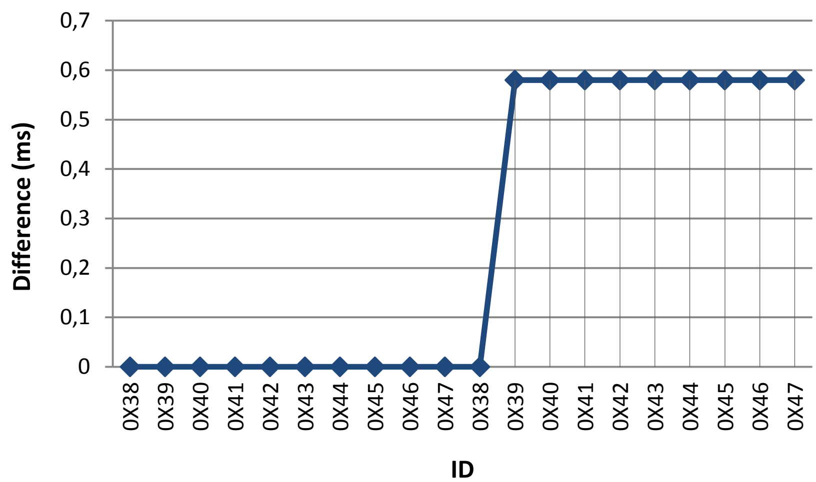


Figure 8. Delay introduced by remote frames on CAN 2

b) Delayed forwarding strategy

In order to guarantee a minimum delay between two consecutive frames of a remote CAN flow on their destination CAN bus, the GW computes, for each decapsulated distant CAN frame, a waiting time since its arrival in the GW to be delayed. We have chosen a waiting time equal to the period of each remote frame. Thus Frames 0X31 and 0X32 have to wait 2 ms at GW2.

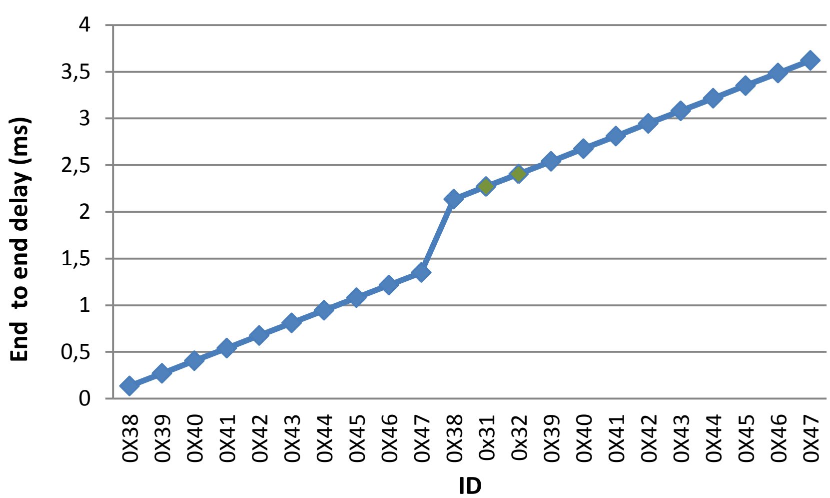


Figure 9. Frames end to end delays (CAN2): local and remote frames

We have presented the end to end delay for local and remote frames in Fig.9. Then, we have compared the delays measured for each type of strategy in the Fig.10, for local frames. We notice that the local delays are lower by using the delayed forwarding strategy. Fig.10 compares the end to end delays of both strategies: immediate forwarding strategy and delayed forwarding strategy. Distant frames are delayed in order to allow the frames to respect their periods at the GW level. Therefore, the worst case delay of the local flows will not augment.

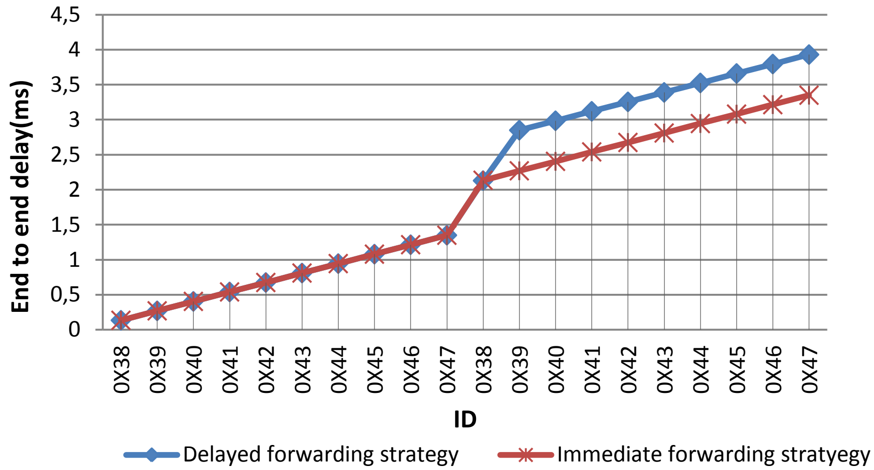
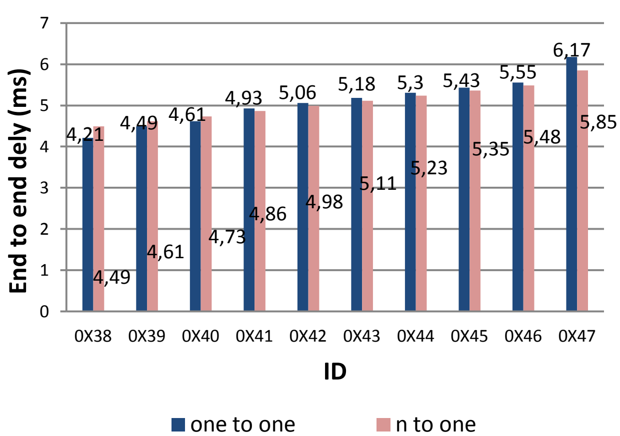


Figure 10. Immediate vs Delayed forwarding strategy

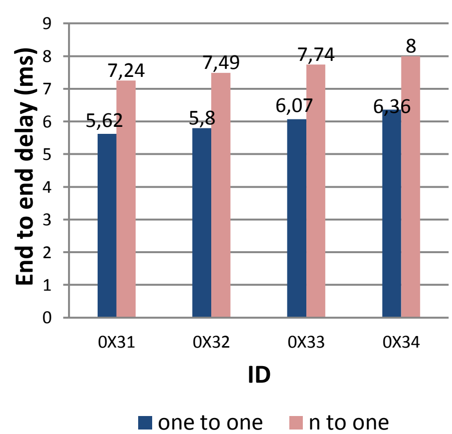
2) n to one bridging strategy

In the previous section, we judge that the one-to-one strategy is sufficient if there is only one remote frame generated by each bus. However, in case of several remote frames, this strategy is no longer efficient and generates a large overhead on Ethernet (Ethernet frames with few data). As a solution, we have proposed the n to one bridging strategy: the GW encapsulates exactly a specific number n of remote CAN frames in each Ethernet frame. This strategy will engender a reduction of Ethernet frames numbers and an increase of their size. Indeed, the use of Ethernet bandwidth is improved. Nevertheless, we realize that this strategy leads to the creation of a waiting delay at the GW for remote CAN frames during encapsulation of the n frames.

We have chosen to compare in Fig.11 the n to one bridging strategy with the previous one to one bridging strategy. So, we consider a new topology with four remote frames. Frames generated on CAN bus1 have 0x31 and 0x33 as identifiers and those of CAN bus2 have 0x32 and 0x34. In the n to one strategy, we establish n equal to 4 and then we encapsulate 4 of remote CAN frames in each Ethernet frame. In this situation, both the one to one and the n to on strategies are tested. Comparison reveals that a delay has been occurred in the remote frames.



(a) Local frames



(b) Remote frames

Figure 11. One to one bridging strategy vs n to one bridging strategy

VI. CONCLUSION

In this paper, we have examined the use of Ethernet in conjunction with CAN for communications in a real-time system. The interconnection between both of them is realized by GWs. We have demonstrated that the bridging strategy implemented at these GWs is a key issue in such architecture. The GW mapping strategy according to its function affects the duration of the message latency at the GW. Therefore, we have studied different CAN/Ethernet bridging strategies and compared their corresponding performance. We have showed that a good strategy consists in adding a specific time on GWs to delay remote CAN frames until their encapsulation in the Ethernet frame (the delayed forwarding strategy).

We are currently working on the evaluation of another heterogeneous network case study using the AFDX avionic network with CAN buses. Moreover, the optimization of an avionic GW could be considered to improve the avionic network real-time performance.

REFERENCES

[1] A. Kern, "Ethernet and IP for Automotive E/E-Architectures," PhD Thesis. Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany, 2012.

[2] J.-L. Scharbarg and C. Fraboul, "A generic simulation model for end-to-end delays evaluation on an avionics switched Ethernet," in *Fieldbuses and Networks in Industrial and Embedded Systems*, 2007, pp. 159-166.

[3] G. Cena and A. Valenzano, "FastCAN: a high-performance enhanced CAN-like network," *Industrial Electronics, IEEE Transactions on,* vol. 47, pp. 951-963, 2000.

[4] R. Makowitz and C. Temple, "FlexRay-a communication network for automotive control systems," in *2006 IEEE International Workshop on Factory Communication Systems*, 2006, pp. 207-212.

[5] A. E. E. Committee, "ARINC Report 655: Remote Data Concentrator (RDC) generic description," *Annapolis, Maryland,* 1999.

[6] A. Varga and R. Hornig, "An overview of the OMNeT++ simulation environment," in *Proceedings of the 1st international conference on Simulation tools and techniques for communications, networks and systems & workshops*, 2008, p. 60.

[7] H. Butz, "Open integrated modular avionic (ima): State of the art and future development road map at airbus deutschland," *Signal,* vol. 10, p. 1000, 2010.

[8] J.-B. Itier, "A380 integrated modular avionics," in *Proceedings of the ARTIST2 meeting on integrated modular avionics*, 2007, pp. 72-75.

[9] A. Radio, "Inc. ARINC Specification 653 Avion ics application software standard interface," ed: Annapolis: Aeronautical Radio, Inc, 1997.

[10] Y. Yuna and H.-g. XIONG, "A method for bounding AFDX frame delays by network calculus [J]," *Electronics Optics & Control,* vol. 9, p. 014, 2008.

[11] R. Queck, "Analysis of ethernet avb for automotive networks using network calculus," in *Vehicular Electronics and Safety (ICVES), 2012 IEEE International Conference on*, 2012, pp. 61-67.

[12] S. Kerschbaum*, et al.*, "Network Calculus: Application to an Industrial Automation Network," *MMB & DFT 2012,* p. 9, 2012.

[13] H. Bauer*, et al.*, "Applying and optimizing trajectory approach for performance evaluation of AFDX avionics network," in *Emerging Technologies & Factory Automation, 2009. ETFA 2009. IEEE Conference on*, 2009, pp. 1-8.

[14] S. Bensalem*, et al.*, "Statistical model checking qos properties of systems with sbip," in *Leveraging Applications of Formal Methods, Verification and Validation. Technologies for Mastering Change*, ed: Springer, 2012, pp. 327-341.

[15] J.-L. Scharbarg and C. Fraboul, "Simulation for end-to-end delays distribution on a switched ethernet," in *Emerging Technologies and Factory Automation, 2007. ETFA. IEEE Conference on*, 2007, pp. 1092-1099.

[16] J.-Y. Le Boudec, "Application of network calculus to guaranteed service networks," *Information Theory, IEEE Transactions on,* vol. 44, pp. 1087-1096, 1998.

[17] Y. Jiang and Y. Liu, *Stochastic network calculus* vol. 1: Springer, 2008.

[18] A. Burchard*, et al.*, "A min-plus calculus for end-to-end statistical service guarantees," *Information Theory, IEEE Transactions on,* vol. 52, pp. 4105-4114, 2006.

[19] H. Bauer*, et al.*, "Improving the worst-case delay analysis of an AFDX network using an optimized trajectory approach," *Industrial Informatics, IEEE Transactions on,* vol. 6, pp. 521-533, 2010.

[20] S. Limal, "Architectures de contrôle-commande redondantes à base d'Ethernet Industriel: modélisation et validation par model-checking temporisé," École normale supérieure de Cachan-ENS Cachan, 2009.

[21] J. Matsumura*, et al.*, "A Simulation Environment based on OMNeT++ for Automotive CAN–Ethernet Networks," *Analysis Tools and Methodologies for Embedded and Real-time Systems,* p. 1, 2013.

[22] J. Hao*, et al.*, "Modeling and simulation of CAN network based on OPNET," in *Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference on*, 2011, pp. 577-581.

[23] A. Kern*, et al.*, "Gateway strategies for embedding of automotive can-frames into ethernet-packets and vice versa," in *Architecture of Computing Systems-ARCS 2011*, ed: Springer, 2011, pp. 259-270.

[24] A. A. Nacer*, et al.*, "Strategies for the interconnection of CAN buses through an Ethernet switch," in *Industrial Embedded Systems (SIES), 2013 8th IEEE International Symposium on*, 2013, pp. 77-80.

[25] D. Thiele*, et al.*, "Formal timing analysis of CAN-to-Ethernet gateway strategies in automotive networks," *Real-time systems,* vol. 52, pp. 88-112, 2016.

[26] J.-L. Scharbarg, *et al.*, "Interconnecting can busses via an ethernet backbone," in *Fieldbus Systems and their Applications*, 2005, pp. 206-213.
[27] H. Ayed, *et al.*, "Gateway Optimization for an Heterogeneous avionics network AFDX-CAN," in *IEEE Real-Time Systems Symposium (RTSS)*, 2011.
[28] J. Jasperneite, *et al.*, "Deterministic real-time communication with switched Ethernet," in *Proceedings of the 4th IEEE International Workshop on Factory Communication Systems*, 2002, pp. 11-18.
[29] G. Embedded, "AFDX/ARINC 664 protocol tutorial," 2007.
[30] R. Knueppel, "Standardization of CAN networks for airborne use through ARINC 825," *Airbus Operations GmbH,* 2012.
[31] R. I. Davis, *et al.*, "Controller Area Network (CAN) schedulability analysis: Refuted, revisited and revised," *Real-time systems,* vol. 35, pp. 239-272, 2007.
[32] I. Standard, "11898: Road Vehicles—Interchange of Digital Information—Controller Area Network (CAN) for High-Speed Communication," *International Standards Organization, Switzerland,* 1993.
[33] N. Rejeb, *et al.*, "Modeling of a heterogeneous AFDX-CAN network gateway," in *Computer Applications & Research (WSCAR), 2014 World Symposium on*, 2014, pp. 1-6.

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 3, March 2016

Reusability Quality Attributes and Metrics of SaaS from Perspective of Business and Provider

Areeg Samir[1] Nagy Ramadan Darwish[2]

***Abstract*-Software as a Service (SaaS) is defined as a software delivered as a service. SaaS can be seen as a complex solution, aiming at satisfying tenants requirements during runtime. Such requirements can be achieved by providing a modifiable and reusable SaaS to fulfill different needs of tenants. The success of a solution not only depends on how good it achieves the requirements of users but also on modifies and reuses provider's services. Thus, providing reusable SaaS, identifying the effectiveness of reusability and specifying the imprint of customization on the reusability of application still need more enhancements. To tackle these concerns, this paper explores the common SaaS reusability quality attributes and extracts the critical SaaS reusability attributes based on provider side and business value. Moreover, it identifies a set of metrics to each critical quality attribute of SaaS reusability. Critical attributes and their measurements are presented to be a guideline for providers and to emphasize the business side.**

***Index Terms*-Software as a Service (SaaS), Quality of Service (QoS), Quality attributes, Metrics, Reusability, Customization, Critical attributes, Business, Provider.**

I. INTRODUCTION

Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly delivered with a minimal management effort or service provider interaction [1]. Software as a Service (SaaS) is a software delivery model in which software resources are remotely accessed by clients [2]. The SaaS delivery model is focused on bringing down the cost by offering the same instance of an application to as many customers, i.e. supporting multi-tenants. Multi-tenancy is one of the most important concepts for any SaaS application.

SaaS users can customize their provider application services so that the services can meet their customers' specific needs [3]. Thus, better customizability will lead to an application with better reusability, which helps in better understanding and lower maintenance efforts for the application. Therefore, it is necessary to estimate the reusability of services before integrating them into the system [4]. Architecturally, SaaS applications are largely similar to other applications built using service-oriented design principles [3]. A service-oriented architecture (SOA) is a collection of services communicates with each other by means of passing data. SOA configures entities to maximize loose coupling and reuse [5]. Thus, reusability of services can be considered as a key criterion for evaluating the quality of services.

Measuring the reusability of application has been done before, however the used metrics are either unsuitable for services or lack expressiveness. Such as the works in [5-11]. There are several problems with the previous mentioned works. The first problem is many of these metrics require analysis of source codes, these metrics cannot be applied to black-box components such as services. The second problem is that most metrics handle reuse instead of reusability and consider them two different concepts. For example, the works in [12 and 13] defined reuse as a way of reusing existing software artifacts or knowledge to create new software while reusability is the degree to which a thing can be reused. The third problem is that some works such as [14] handled reusability from the service consumers' point of view and neglected the provider and business sides. Finally, current works [4, 5, and 10] on evaluating reusability in SOA are mostly on service components not on services.

The contributions of this paper are first, the most common quality attributes of SaaS applications based on provider and business sides will be introduced. Second, the critical quality attributes of SaaS based on reusability and customizability will be derived. These attributes will work as a guideline to the providers in order to aid them delivering reusable and customizable SaaS. Third, presenting a metric suite to measure the derived quality attributes and to evaluate the effectiveness of tenants' customization on the reusability of services. Fourth, improving the Service Measurement Index (SMI), which is cloud services measurement standards, through expanding SMI framework with the proposed critical attributes and metrics. Reusability and customizability have been chosen because they are often regarded as essentials to service design and quality.

The remainder of the paper is organized as follows. Section II presents background material on Reusability, Customizability, Quality of Service, SMI, and ISO. Section III depicts the research methodology of extracting critical attributes of reusability of SaaS reusability. Section IV shows the common SaaS quality attributes of providers and business. Section V provides the critical SaaS reusability quality attributes and their metrics. Sections VI gives an evaluation with the related work. Finally, section VII provides a conclusion.

II. BACKGROUND

[1]Information Technology and System, Institute of Statistical Studies and Research
[2]Information Technology and System, Institute of Statistical Studies and Research

This section will provide a background knowledge about the reusability, customizability, quality of service, and Service Measurement Index.

A. Reusability

Reusability is the basic concept of software engineering research and practice, as a mean to reduce development costs, time, improve quality and component based development [15]. Reuse refers to using components of one product to facilitate the development of a different product with a different functionality [16].

There are three types of software reuse *black box*, *white box*, and *glass box*. Choosing the reusability types are based upon software testing development. For example, selecting the white box allows the programmer to share the internal structure of the box with others through inheritance unlike the black box, which prevents the internal structure from sharing. The glass type allows the inside as well as the outside of the box from being seen. Software reuse can apply to any life cycle product, not only to fragments of source code [15].

Reusability research literatures such as [17, 18, 19, and 20] pointed out that there are many issues involved in software reuse should be addressed. First, which part can be reused? Second, how the reusable part can maximize the economic value. Third, how to adapt the reusable part to the needs of a wide variety of tenants. Fourth, what is the best quality model to evaluate SaaS customizability and its imprints on reusability? Finally, what are the proper metrics that can be used to measure the effectiveness of quality model?

B. Customization

Customization expresses the "*modification of packaged software to meet individual requirements*" [21]. SaaS application provider offered an application template to achieve the tenant customization. The application templates have unspecified places "Customization Points" that can be customized to fit each tenant requirements [22]. Customization can take place at any layer of SaaS application layers such as GUI, Process, Service, or Data layer.

Customizability plays a special role in reuse. Reusing in SaaS applications requires changing them to suit the new requirements of tenants. If tenants cannot do their modifications easily, it indicates that the code is less likely to be reused. Thus, measuring customization quality, evaluating its effect on reusability of SaaS applications, enhancing the adaptability of customization, and improving its understandability are considering important tasks that need to be addressed in SaaS applications.

C. Quality of Service

Quality describes to which degree a system meets specified requirements. Quality attribute is a characteristic that affects the quality of software systems [23].

Quality of Service (QoS) is a non-functional component, which can be defined as the ability to provide different priority to different applications, users, data flows or to guarantee a certain level of performance. SaaS needs QoS because it is a crucial factor for the success of cloud computing and it should be delivered as expected to save provider's reputation [24].

To evaluate a service quality, a quality model should be defined. The quality model consists of several quality attributes that are used as a checklist for determining service quality. Quality attributes contain set of measurable service properties such as performance, understandability, availability…etc. In addition, they are often composed of sub-attributes, which evaluate the effectiveness of the parent attribute. Some papers such as [25 and 26] refer to the quality attributes as a Service Level Objects (SLOs) that are considered a core item for Service Level Agreements (SLA) between provider and customer.

The quality model is dependent of the type of service and the one can either use a fixed already defined quality model or define his/her own. Moreover, there are several quality models. Each model differs in its attributes such as McCall's addressed the reusability attribute while ISO/IEC 9126 didn't handle it [27] and SMI neither handled reusability nor customizability [28].

Achieving quality attributes must be considered throughout design, implementation, and deployment. Within complex systems, quality attributes can never be achieved in isolation. The achievement of any attribute will have a positive or negative effect on the achievement of others [29].

D. Service Measurement Index

The Service Measurement Index (SMI) is a new cloud computing standard method based on ISO and developed by the Cloud Services Measurement Initiative Consortium (CSMIC). CSMIC is introduced by Carnegie Mellon University through presenting two SMI versions [30]. SMI defined Key Performance Indicators (KPI) to measure the cloud services quality and performance depending on critical business and technical requirements of industry and government customers [28]. The new version of SMI is the 2.1. It consists of seven categories each category contains several attributes. Each attribute has KPIs to be used in measuring the providers' efficiency. The top-level categories of the SMI framework include Accountability, Agility, Assurance, Performance, Financial, Security and Privacy, Usability. Fig 1 shows the SMI categories and their attributes.

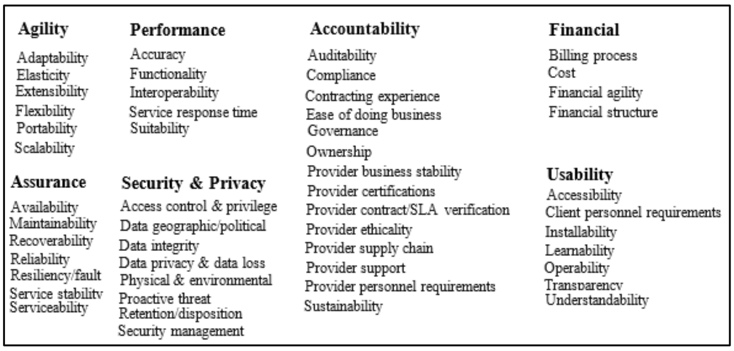


Fig1. SMI Framework

As shown in the previous figure, SMI framework does not consider reusability or customizability of SaaS. Thus, the framework needs to be enhanced in order to tackle those categories.

E. ISO 9126/25010

The ISO 9126 is part of the ISO 9000 standard, which is the most important standard for quality assurance. It identifies six main quality characteristics, namely Functionality, Reliability, Usability, Efficiency, Maintainability, and Portability. These characteristics are broken down into sub-characteristics. Fig 2 demonstrates the ISO 9126 [31]. ISO9126 did not clarifies reusability as one of its attributes. Moreover, it mentioned changeability without specifying a definition or mentioning its functionality.

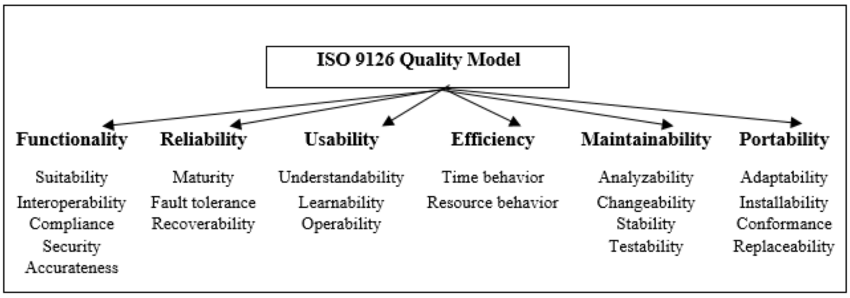


Fig2. ISO9126 Quality Model

ISO 25010 defines System and software quality models characteristics. It considered a successor of ISO9126 [32]. The quality model of 25010 comprises eight quality characteristics in contrast to ISO9126. Each characteristic has sub-characteristics as shown in Fig 3. ISO 25010 classified reusability attribute as sub-attributes of the maintainability attributes but it did not mention the customizability attribute.

Fig3. ISO25010 Quality Model [32]

III. EXTRACTING CRITICAL QUALITY ATTRIBUTES OF SAAS

This section present the steps of extracting SaaS critical attributes of reusability. As shown in Fig 4, the extraction methodology passed through four Stages, which are:

- Stage one studying the quality attributes of SaaS, Service, and Software.
- Stage two identifying common quality attributes of SaaS, Service, and Software.
- Stage three presenting the quality attributes of reusability of SaaS.
- Stage four extracting SaaS critical quality attributes of SaaS.
- Stage five identifying and proposing metrics related to critical attributes of reusability.

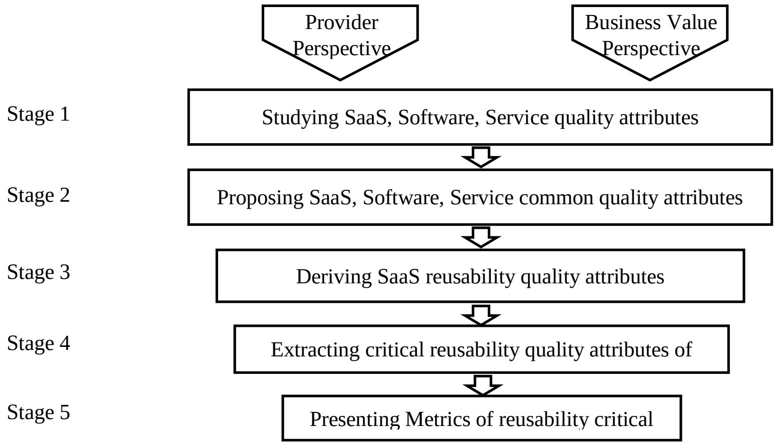


Fig4. Methodology of Extracting Reusability Critical attributes of SaaS

The extraction methodology of five stages is based on the attributes that targeting provider and business value perspectives. The figure demonstrates that starting from the top, the base of pyramid contains many attributes, and by inspecting common attributes, the scope of attributes shrinks gradually until obtaining the critical attributes of SaaS reusability. As soon as the process of extracting critical attributes completed, the phase of proposing and identifying metrics of SaaS critical quality attributes of reusability starts to depict the effectiveness of each critical reusable attribute of SaaS.

IV. SAAS PROVIDER AND BUSINESS QUALITY MODEL

According to the literature studies [7, 19, 24, and 33-38], quality attributes can be classified into four categories, which are provider side attributes, developer side attributes, customer side attributes, and business side attributes. However, the quality attributes for Services and SaaS differ in identifying the quality attributes specifically for reusability.

Reusability has several quality attributes these attributes are not fixed and they are differed from a research work to another. Specifying these attributes depending on the organizations needs' and relating them to a suitable type of measurements are a crucial task. Quality attributes are considered crucial attributes to achieve the provider business objectives. The following sub-sections will provide the most common SaaS quality attributes based on current literature works according to provider and business sides. In addition, the related attributes of reusability, customizability and the overlapped attributes among them will be presented.

A. Provider Quality Attributes Classification

The provider quality attributes are the attributes that related to service providers, which improves their offered services. Sometimes as depicted in [38] the provider attributes are called "*Strategy Specific Attributes*".

Table1 divided into three major columns. First, the *Common 3S Attributes* column which presents the most common quality attributes of SaaS , Software, and Service that have been extracted from studying the existed quality models [4, 7, 14, 17, 19, 24, 33, and 35-44], SMI, and ISO 1926/25030. Second, *the current research works* column, shows the common quality attributes for provider based on SaaS, Service, and Software that have been addressed by the previous mentioned researches. Third, the *Reusability/Customizability Attributes*, which demonstrates the common quality attributes of reusability and customizability in SaaS, Service, and Software separately. The quality attributes of reusability and customizability have been divided into attributes and sub-attributes. Therefore:

- 'R' and 'C' refer to the sub-attributes of reusability and customizability attributes sequentially.

- 'R+' and 'C+' refer to the reusability attribute and customizability attribute.
- The overlap between reusability and customizability sub-attributes might be happened in the same column. Thus, it will have the 'RC' as a metaphor such as variability, commonality, and adaptability. The 'R+ C' means that customizability is a sub-attributes of the reusability attribute (parent-child relation). The 'R C+' illustrates that the R means the customizability attribute is a child of the reusability parent (reference to reusability) and C+ defines that the customizability is clarified as an attribute for some related works.
- Each element 'R' or 'C' in the *Reusability/Customizability Attributes* column points to the sub-attributes it achieves in the *Common 3S Attributes* column.
- The Reusability 'R+' and Customizability 'C+' attributes are highlighted to connect their implemented works in *Current research works* column with Rs and Cs. For example, the Rs of SaaS column mean that the Adaptability, Commonality, Composability, Functionality, Generality, Nonfunctionality, Understandability, and Variability are considered sub-attributes of their parent reusability attribute 'R+' and the works in 'R+', which are [14], [35], [41], [37], are the works that implemented the Rs. Definitely, not all works mentioned all the sub-attributes. For instance, the work in [14] only addressed SaaS reusability sub-attributes, which are the Adaptability, Composability, Generality, and Understandability that are marked in bold. The same as for the work in [35] and [45]. While [41] mentioned the reusability, it did not provide sub-attributes.

The following Table shows that, there are three works [4] [42], and [46] mentioned customizability as an attribute without specifying sub-attributes and four works [47-50] classified customizability as a sub-attributes of reusability. The work in [47] defined customizability as a child of adaptability, which is a child of reusability. For SaaS customization just three works handled it which are [42], [51], and [46]. Moreover, Table I depicted that four works handled reusability attribute [14], [35], [41], and [37]. The work in [4] noted customization according to software component while [42] illustrated customization in SaaS. Regarding reusability, only two works in [14] and [35] explicitly addressed reusability in SaaS unlike [41] and [37] that identified reusability in service design and software sequentially. Therefore, according to the works in Table I and to the literature studies, there is a need to provide a quality model that classifies the critical reusability attributes of SaaS applications and outlines the effect of customization on the overall reuse of SaaS application.

TABLE I
SAAS, SERVICE AND SOFTWARE COMMON QUALITY ATTRIBUTES AND REUSABILITY AND CUSTOMIZABILITY ATTRIBUTES FOR PROVIDER

Common 3S Attributes	**Current research works**	**Reusability/Customizability Attributes**		
		SaaS	**Software**	**Service**
Adaptability	[7], **[14]**, [24], [36], **[37]**, [42], [47]	R	R C	R
Autonomous	[17]			R
Availability	[7], [19], [24], [33], [40], [35], [36], **[37]**, [42], [38]		R	R
Commonality	[7], [44], [51]	R C	R	R
Complexity	**[37]**		R	
Composability	[4], **[14]**, [17], [19], [24], [41]	R	R	R
Configurability	[19], [45], [51]	C	R	
Conformance	[7], [36]			R
Customizability	[4], [42], [47], [51], [46], [48], [52], [53], [49], [50]	C+	R C+	
Discoverability	[7], [17]			R
Efficiency	[24], [35], **[37]**, [53]		R	
Extensibility	[24], **[37]**, [42], [38]		R	
Flexibility	[24], [36], [43], [38], [45], [51]	C		R
Functionality	[17], [24], [40], **[35]**, [41] , [42], [45]	R	R	R
Generality	**[14]**, [17], **[37]**	R	R	R
Granularity	[17]			R
Integrity	[4], [24], [33], [42]		R	
Interoperability	[19], [33], [40]			
Coupling-ability	[41], **[37]**, **[45]**	R	R	R
Maintainability	[24], [40], **[37]**, [43], [38], [45]		R	
Modifiability	[33]			
Modularity	[7], **[37]**, **[45]**	R	R	R
Multitenancy	[19]			
Nonfunctionality	[7], **[35]**, [41]	R		
Performance	[19], [33], [40], [42], [38]			
Portability	[4], [17], [19], [40], **[37]**, [42], [43], [38], [45]		R	R
Recoverability	[24], [38], [45]			
Relevance	[28], [40]			R
Reliability	[17], [24], [33], [40], [35], [36], **[37]**, [42], [38], [45]		R	R
Resiliency	[19], [24], [40], [42]			

Reusability	[14], [35], [41], [37], [45], [47], [48], [49], [50]	R+	R+ C	R+
Scalability	[24], [33], [35], [42], [38]			
Security	[19], [24], [33], [40], [36], [42], [38], [45]			R
Stability	[24], [42], [43], [45]			
Statelessness	[17]			R
Understandability	[4], **[14]**, [17], [**37**], [43]	R	R	R
Usability	[19], [33], [40], [42], [38], [45]		R	
Validability	[33], [40] , **[37]**, [45]		R	
Variability	**[35]**, [43], [44], [51], [52], [53]	R C	R C	

B. Business Quality Attributes Classification

Business quality attributes are attributes that imprinted organization business [38]. The quality attributes of the existing quality models for SaaS from the business side were reviewed. The quality attributes have been specified 37 business attributes for SaaS. Each attribute contains set of attributes. The attributes are extremely varying from one work to another. Thus, for simplicity, only the upper level of attributes has been considered. Table II displays the popular SaaS business quality attributes that have been addressed by existing literature studies [38] and [54-69].

TABLE II
SAAS BUSINESS ATTRIBUTES

Common Attributes	**Current research works**
Accessability	[62], [67]
Adaptability	[54], [58], [66]
Agility	[54], [67], [69]
Availability	[38], [54], [58], [60], [61], [62], [66]
Cash flow	[57], [59], [65]
Churn	[57], [59]
Conformance	[61]
Continuity	[58]
Cost	[57], [58], [62], [63], [64], [66], [67], [69]
Deployment	[63]
Effectiveness	[59]
Efficiency	[55], [58]
Extensibility	[54]
Flexibility	[67]
Functionality	[55], [58]
Integration	[64], [67]
Interoperability	[54], [66]
Maintainability	[55]
Modifiability	[54]
Performance	[38], [54], [69]
Portability	[55]
Productivity	[61], [68], [69]
Profitability	[59], [61], [65]
Quality	[58], [60], [61], [67], [69]
Recoverability	[38], [59], [68]
Reliability	[38], [54], [55], [58], [60], [66]
ROI	[58], [59], [61], [63]
Revenue	[56], [57], [59], [67]
Risk	[57], [58], [61], [67]
Scalability	[54], [60], [62], [67], [68]
Security	[54], [58], [60], [64], [67], [68]
Suitability	[64], [66]
Support	[60]
Sustainability	[58]
Testability	[54], [57]
Upgradability	[57], [63], [67], [68]
Usability	[55], [58], [61]

As illustrated in Table II, most of the works focused on Security, Reliability, Scalability, Quality, Cost, and Availability. Only two works mentioned the Churn, which considered an important attribute in measuring the business effectiveness and customer satisfaction. Moreover, just four works have addressed Return of Investment (ROI), Risk, and Revenue, which considered essential requirements for every SaaS business sector.

According to the previously mentioned literature studies in Table II, no work has specified the attributes that imprint reusability and customizability in business. Consequently, a quality model that explicitly clarifies the critical attributes of SaaS from the business side based on reusability and customization is required.

C. Reusability Quality Model of Provider and Business

According to the quality attributes of provider and business and depending on the existed quality models, the reusability quality attributes for SaaS quality model of these two concerns will be derived. The reusability attributes that relate to provider take the "P" symbol and the attributes that belong to business will have the "B" symbol as depicted in Fig4. The intersection represents the attributes that are related to business and important to provider as well. Fig 5 demonstrates customizability is considered one of the reusability sub-attributes. It has a direct effect on the reusability. The highly customized SaaS application the more reusable it could be. Fig 6 illustrates the customizability sub-attributes that have been chosen depending on the current literature studies and the existed models in Table I.

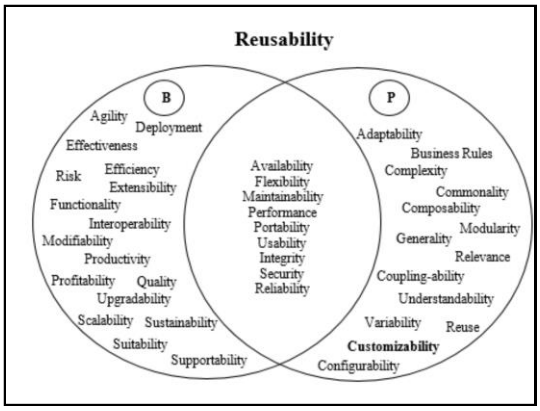


Fig5. Reusability Quality Attributes

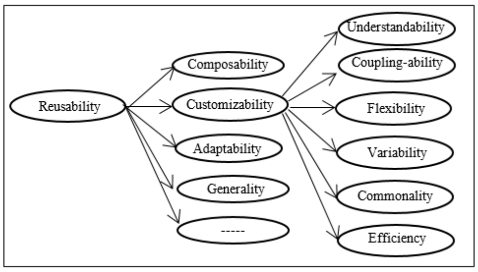


Fig6. Customizability Quality sub-attributes

As depicted in Fig 6, customizability has an effect on reusability. It has six sub-attributes, which are understandability, flexibility, variability, commonality, efficiency, and coupling-ability. Each sub-attributes will be explained and measured in the following section.

The interaction between SaaS layers, provider attributes and business attributes are shown in Fig 7. The providers offer SaaS considering its three main layers, which are Presentation layer (GUI), Business Process layer (Business Logic) and its composite services, and Data layer (Data Logic). Both the provider attributes and business attributes have a direct effect on SaaS layer. The more a provided SaaS supports and enriches provider and business attributes, the more customers will use a software. Each layer component of SaaS application can be customized and reused in other applications.

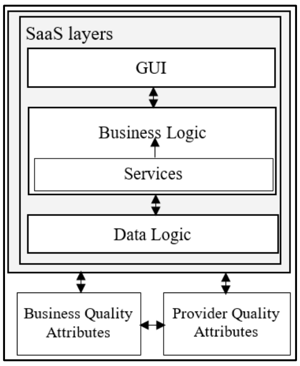


Fig7. The interaction among SaaS layers, provider attributes, and business attributes

The following is a description of each layer.

- SaaS Layers:
  - SaaS Presentation layer –the graphical user interface such as browser that is used to provide a software access.
  - SaaS Business layer–the business logic layer that contains the business domain activities, rules, tasks, and constrains. It composed related services to provide a workflow that satisfies clients' requirements.
  - SaaS Data layer–the data logic layer responsible for retrieving stored data from a storage database or file system.
- Provider Quality Attributes –the quality attributes which relate to provider such as reusability, understandability…etc. it effects each layer of SaaS applications. Achieving the quality attributes allow providers to produce a high quality software that acquires many customers and increases provider profitability.
- Business Quality Attributes –the quality attributes that target the business value and effect organization business such as revenue, churn, availability…etc. Enhancing the business attributes has a significant impact on provider.

V. THE CRITICAL QUALITY REUSABILITY ATTRIBUTES AND MEASURES

The previous section covered a wide range of quality attributes for SaaS, software, and service. Around 39 quality attributes have been found in the abovementioned literatures targeting the provider, and 37 quality attributes found in the literature for business side. However, there is no unified definition or complete list of quality attributes for SaaS provider and business. Moreover, there is no single universal classification model that accommodates all the critical quality attributes that needed for different types of SaaS applications.

Reusability and Customizability quality attributes of SaaS have been derived depending on SaaS common features, current models on SaaS quality attributes, SMI standard, and ISO25010. In addition, the SaaS attributes are not only targeting the reusability but it also exploring customization effect on reuse.

A. The Reusability Quality Model

This paper has chosen the following quality attributes to be classified as the critical attributes for provider and business of SaaS applications based on reusability. Each attribute will be defined and will be measured to compute the values of the attributes. The provider and business reusability attributes have been classified into critical, basic, and optional attributes as shown in Table III.

- The critical attributes means that attributes must be existed to provide a reusable SaaS.
- The basic attributes clarifies that attributes should be listed to achieve the critical attributes depending on provider application.
- The optional attributes means that attributes may or may not be included in SaaS applications.

For example, maintainability, agility, usability, financial, security, and performance are considered a critical attributes. Each critical attribute has sub-attributes such as understandability, modularity, customizability, reliability, functionality…etc.

TABLE III
CRITICAL, BASIC, AND OPTIONAL REUSABILITY ATTRIBUTES OF SAAS

Attributes	**Sub-attributes**	
Critical	**Basic**	**Optional**
Maintainability	Understandability Modularity Composability	Reliability Availability

	Customizability Generality Complexity	
Agility	Adaptability Relevance Portability Extensibility	--
Usability	Configurability Efficiency	--
Financial	Profitability	Upgradability
Performance	Functionality	Service Response Time Execution Time
Security	Integrity	--

Reusability has indirect effect on the critical attributes. For example, in order to enhance reusability of SaaS application, the application should be more understandable and customizable as a result the maintainability of the application improved. Unlike the basic and optional attributes which have direct imprint on SaaS reusability. For instance, to provide a reusable application it should be application independent, can be used on any platform or programming language, and should achieve high revenue.

Each attributes and sub-attributes have an effect on each other. For example, providing an understandable reusable application will improve maintainability and usability of the application. Moreover, good customizable application not only enhances maintainability but it also allows SaaS application to be more adaptable. In addition, reusability and customizability have influence on security attribute. Flexibility attribute has an impact on customizability, configurability, composability. The definition of each attribute and sub-attribute will be given in the following table. In addition, the critical attributes (attributes) are marked in bold whereas the Basic and optional attributes (sub-attributes) are listed for each attributes.

TABLE IV
THE REUSABILITY QUALITY ATTRIBUTES DEFINITIONS

Maintainability	Measures the ability of SaaS provider modifications to provide a service in a good condition
Understandability	Specifies how simple a service capabilities and functions can be understood
Modularity	Measures the ability of a service to provide independent functionality without relying on other service. Well modularize service leads to a good reusable service.
Composability	The ability of a service to be composed for achieving tenants requirements
Customizability	Measures the ability of service provider to modify services for meeting tenants' requirements. Highly customized SaaS services will produce a reusable SaaS that satisfy multitenant different needs
Generality	Depicts the ability of the reusable service to be generic for achieving current and upcoming tenants requirements'
Complexity	Measures the simplicity of SaaS services to be reused with other services. Complex service decreases its ability from being reused with other SaaS services.
Reliability	Measures the amount of time that SaaS keeps operating after reusing some services without indicating failure
Availability	Indicates the uptime service availability after applying the reusability. Services with low availability would influence a negative impact on business and provider reputation
Agility	Measures the service impacts on tenants modification with minimum confusing
Adaptability	Measures the ability of a service to be adapted to achieve multitenant requirements
Relevance	Demonstrates to which extent a reusable service is relevant with the rest of SaaS service
Extensibility	Clarifies the ability to add features (reusable parts) to new or existed SaaS services
Usability	Measures to which extent SaaS service can be used, configured, and executed, when it is used in certain conditions
Configurability	Measures the degree of configuration provided by provider such as configuring presentation layer, data layer, or business logic layer. Providers who design highly configuring SaaS will result a SaaS service that support reusability
Efficiency	Measure the efficiency of reusing resources of SaaS service to conduct its function
Financial	The amount of money gathered or spent by providers
Profitability	Measures provider revenue, the cost to acquire and retain customers
Cost	Measures the cost of service reuse with modification and without modification
Upgradability	Specifies how well upgrading SaaS with reusable service(s) imprinting provider revenue
Performance	Measure the SaaS reusability performance
Response Time	Measures the amount of time taken to respond after reusing
Execution Time	Measures the amount of time taken to execute SaaS services after reusing
Security	Measures the effectiveness of SaaS providers controls on using reusable services with other SaaS services
Integrity	Indicates that using a reused service with other SaaS service follow SaaS service security roles

B. Reusability Metrics

Metrics are used for measuring the quality and performance of a specific unit such as service, software, or components...etc. In order to realize the effectiveness of SaaS reusability, the three classifications of SaaS reusability will be measured, which are the critical, basic, and optional attributes. The following subsections will explore the existed reusability metrics of SaaS and Services and will provide measurements for the provider and business quality model mentioned in section III part A.

1) Existed Reusability Metrics

Multiple metrics and methods have been proposed by many authors. According to the performed literature review, 11 papers were found measuring reusability. Around four papers measured reusability of service [7, 41, 70, and 71], one paper measured reusability of cloud services in general [14], and two papers measured SaaS reusability [35 and 44]. Some SaaS reusability papers provided a direct measure [14 and 35], other measured reusability indirectly [44]. According to the literature review, the reusability metrics that targeted SaaS and Services have been extracted as depicted in Table V. The service oriented metrics have been chosen because service oriented is considered the building block of many SaaS applications.

TABLE V
REUSABILITY METRICS

Reusability Attributes	Reusability Metrics	Papers Ref.	Disadvantages	Domain
Understandability, Publicity, Adaptability, Composability	Comprehensibility of Service (CoS), Awarability of Service (AoS), Coverage of Variability (CoV) and Completeness of Variant Set (CoA), Modularity of service (MoS), Interoperability of Service (IoS)	14	Neglecting Business side Adaptability has specified as a customizability in measurement	SaaS
Reusability	Functional Commonality (FC), Non-functional Commonality (NFC), Coverage of Variability (CoV)	35	Neglecting Business side Not considering reusability attributes Specifying variability as a metric of reusability	SaaS
Commonality, Variability	Commonality, Variability	44	Neglecting Business side Not considering other reusability attributes Specifying steps as a measure for variability	SaaS
Business Commonality, Modularity, Adaptability, Standard Conformance, Discoverability	Functional Commonality (FC), Non-Functional Commonality (NFC), Modularity, Adaptability, Standard Conformance (SC), Syntactic Completeness of Service Specification (SynCSS), Semantic Completeness of Service Specification (SemCSS), Discoverability	7	Not considering other reusability attributes Neglecting the other business attributes	Services
Understandability, Adaptability, Flexibility, Portability, Independence, Modularity, Generality	Existence of meta-information (EMI), Rate of Service Observability (RSO), Rate of Service Customizability (RSC), Self – Completed of Service's return Value (SCSr), Self – Completed of Service's parameter (SCSp), Density of Multi-Grained Method (DMG)	70	Not considering reusability attributes Neglecting Business side	Services
Process Reusability	Mismatch Probability (MMPs)	71	Only considered service compatibility and neglecting the other attributes Not considering Business side	Services
Reusability	Service Reuse Index (SRI)	41	Not determining reusability attributes Not considering Business or Provider side	Services

As shown in the preceding table there are 27 reusability metrics that evaluate provider quality models. Three papers focused on measuring SaaS reusability [14, 35, and 44] and four papers measured reusability of services [7, 41, 70, and 71]. In addition, six attributes (*Understandability, Publicity, Adaptability, Composability, Commonality, and Variability*) are used in identified the SaaS reusability metrics and eight attributes (*Commonality, Modularity, Adaptability, Standard Conformance, Discoverability, understandability, Flexibility, Portability, Reusability*) are specified to measure services reusability. The table illustrates the disadvantages of each reusability SaaS and Services metrics.

2) The Proposed Reusability Metrics

Depending on the studies and the extraction steps that have been performed in the previous sections, the reusability metrics for SaaS will be identified and proposed for each sub-attribute of the critical quality attributes to be more understandable.

- **Understandability**
  The paper in [14] provided a metric to measure the reusability of cloud services. The paper specified that the service could not be reused if it is not understandable. The metric used service comprehensibility to measure service understandability.

$$Comprehensibility\ of\ Service\ (CoS) = \frac{Number\ of\ Field\ with\ Acceptable\ Readability}{Total\ Numbers\ of\ Fields}$$

As mentioned by [72] increasing in size and complexity of software have a bad impact on several quality attributes specifically understandability and maintainability.

- **Modularity**
To measure a service independency from other services the paper in [7] divided the number of service operations that are rely on other service by the total service functionalities. To measure the metric result, a value range from 0-10 is used to indicate the service independency.

$$Modularity = \frac{Number\ of\ Dependent\ Service\ Operations}{Total\ Number\ of\ Service\ Operations}$$

The paper in [73] has been chosen three criteria to measure a service modularity, which are understandability, decomposability, and composability. The paper evaluated metric result by giving an absolute scale whose value ranges from 0-10.

  - Decomposability Metric:

$$Operations\ Relationship\ Degree\ (ORD) = \frac{Entities\ Relationship\ Degree}{Number\ of\ Edges\ of\ Graph\ G}$$

  - Understandability Metric:

$$Operations\ Understandability\ (OUD) = \frac{The\ Amount\ of\ Understandability\ of\ Entities}{Business\ \ Understandability\ Amount\ of\ Graph\ G}$$

  - Composability Metric:

$$Service\ Composability\ Degree\ (SCD) = \frac{Business\ Entities\ Composability\ Degree}{Service\ Operation\ Numbers}$$

- **Composability**
It is the ability to combine several services together to get a composed one. The paper in [14] specified two metrics for measuring composability, which are the modularity of service and interoperability of service.

$$Modularity\ of\ Service\ (MoS) = 1 - \frac{Number\ of\ Elements\ with\ External\ Dependencey}{Total\ Number\ of\ Elements}$$

Furthermore, [41] measured composability considering two factors the compositions that a service participate on it, and the number of distinct composition participants.

- **Complexity**
The complexity of a service specifies how maintainable it is and how easy it can be adapted in new service. The complexity of a service can be calculated by the Size of the Service, and Simplicity of Operations. If a service complexity increases, its reuse in other services would be difficult.
- **Generality**
It can be evaluated through developing a reusable service that complies with current requirements of tenant and handles unknown future requirements [17].
- **Customizability**
Providing a customizable SaaS is an important task to meet multiple tenants' requirements. To measure customization there are many sub-attributes that need to be considered to provide affective customization. As mentioned in Fig 5 section III the customization sub-attributes are understandability, variability, commonality, coupling-ability, flexibility, and efficiency.
  - ***Variability*** can be measured by counting the number of variation points that can be customized in the application. In addition, the number of variants for each variation point needs to be measured through divided the number of variants of variation points by the total number of supported variants in application. The papers in [35] measured variation points while [14] supported variation points and variants measurements.

$$Coverage\ of\ Variability\ (CoV) = \frac{Number\ of\ Variation\ Pointd\ Supported}{Total\ Number\ of\ Potential\ Variation\ Points}$$

$$Variant\ Set\ Completeness\ (CoA) = \frac{CoV}{Number\ of\ Variation\ Points}$$

- ***Commonality*** can be defined by specifying common features around the SaaS application. The paper in [36] measured commonality through dividing the number of applications needing high feature commonality by the total number of target applications. Services with high commonality would yield high return on the investment.

$$Customization\ Commonality = \frac{\text{Number of Applications needing High Feature Commonality}}{\text{Total Number of Target Applications}}$$

- ***Coupling-ability*** specifies that the changing in one variation point should not affect the rest of application. As the number of couples becomes larger, the ability to change in application is lower and as a result, customization and maintenance be more complex. Coupling-ability can be measured through calculating the number of relationships between variation points, variants, and variation points and variant in application. The larger the value of coupling-ability metric, the tighter the relationship with other variation points and variants.
- ***Understandability*** provides an application that supports simple and correct customizations to be understood by customizers. It can be achieved by calculating the number of variation points in the application and separating them from the common points in the same application.
- ***Flexibility*** metric provides a flexible SaaS that adapts different requirements of multitenant, which significantly will imprint application customization. Flexibility can be measured by calculating the number of allowed alternative variants for each variation points. As the number of supported alternatives becomes larger, the ability to provide a flexible customizable application is higher. The ability to add or remove customizable parts from SaaS service not only enhances flexibility but it also has a significant impact on maintainability and reusability.

$$Customization\ Flexibility = \frac{Number\ of\ Supported\ Alternative\ Variants}{Total\ Number\ of\ Variants}$$

- ***Efficiency*** metric measures the quality of the overall customization. This can be done through merging all customization metrics and giving a value range to determine the customization efficiency.

- **Reliability**

  It can be calculated considering three metrics.
  - The first metric is about measuring *Recoverability*. This metric check whether a service after applying the reusable
  - The second metric is the *Fault Tolerance*, which indicates whether a service after applying the reusable part can maintain a specified level of performance in case of faults.
  - The third metric is calculating the *Amount of Time Period* that specifies the time, which a service operated after applying the reusable part without indicating faults, and the failure mean time, which calculates the average time between one to another failures. High reliable service indicates high amount of time working without faults and with minimum failure mean.

  The paper in [35] specified two metrics to measure reliability, which are:
  - Service Stability:

$$Coverage\ of\ Fault\ Tolerance\ (CFT) = \frac{Number\ of\ Unfailures\ Faults}{Total\ Number\ of\ Ocurring\ Faults}$$

$$Coverage\ of\ Failure\ Recovery\ (CFR) = \frac{Number\ of\ Filures\ Remedied}{Total\ Number\ of\ \ Failyres}$$

  - Service Accuracy:

$$Service\ Accuracy = \frac{Number\ of\ Correct\ Reponses}{Total\ Number\ of\ requests}$$

  Moreover, reliability can be measured through calculating the probability of service processing in specific time interval [74].

- **Availability**

  The works in [24, 33, and 38] specify that *Uptime* can be used as indicator of service availability.

In [30] *Robustness of Service* measured availability through dividing the available time to invoke SaaS by the total time for operating SaaS.

$$Robustness\ of\ Service = \frac{The\ Available\ Time\ to\ Invoke\ SaaS}{Total\ Time\ for\ Operating\ SaaS}$$

The proposed work has specified Availability Time of Service as a metric to measure the availability of service before and after performing the reusability to check whether a reusable part enables SaaS to perform its functionality in a specific time to the total time expected to function.

- **Adaptability**
  To measure if the reusable service can be adapted in multiple SaaS to achieve multitenant requirements, it is required to satisfy the variable parts and their variants in SaaS. Thus, the higher effective customization, the more adaptable service can be achieved and the easier to modify in SaaS which makes the business more adaptable to the requirements changing.
- **Relevance**
  A service relevance metric estimates services reusability through estimating compositions numbers contained in a service, clustering services into domains, analyzing relations among services, and estimating the potential impact of new services [39]**.**
- **Extensibility**
  The works in [24 and 38] measured extensibility by supporting the ability to handle and add new services or functionality to existing SaaS in the future. Extensibility has an effect on agility and reusability of SaaS. Supporting extensibility allows providers producing quick products to convoy multitenant different requirements.
- **Configurability**
  It can be measured through calculating the number of configurable parameters per layer divided by a Total number supported by a configuration file**.** The effectiveness of the configuration will be determined by specifying a range value from 0 to 10. A high configuration has a considerable impact on usability and reusability of SaaS. High configuration means, a configuration file supports simplicity in changing parameters, has size limit, and defines type**.**

$$Service\ Configurability\ (SCon) = \frac{Number\ of\ Configurable\ Parameters}{Supported\ Total\ Number\ of\ Config\ File}$$

- **Efficiency**
  The paper in [33 and 35] provided measures to evaluate SaaS utilization efficiency. The first paper introducing two measures, which are the resource utilization and the Time behavior. The second paper using time as an indicator of SaaS utilization efficiency. From these measures, the reusability efficiency of SaaS can be determined by using three metrics.
  - First metrics is the Utility of Reusable Service (URS) that measures the Amount of Reusable Services (ARS) divided by the Total number of Predefined Services (TPS).

$$URS = \frac{ARS}{TPS}$$

  - Second metric, calculates the Reuse Saving Time (RST), which measures the amount of time saved by performing reusability, or by the length of the transaction taking after applying the reusable part.

$$RST = TSR - TSWR$$
$$where, TSR\ is\ the\ Time\ Saved\ without\ Reusabilit$$
$$TSWR\ is\ the\ Tine\ Saved\ With\ Reusability$$

  - Third metric, Reusability Indicators (RI) that works by giving an index to reusable services to specify the most reusable used service. The indicators specifies range numbers from 0 that refers to low reuse, until 10, which reflect high reuse.

  The values of the URS, RST, and RI metrics will be evaluated according to a range value from 0-10. A higher result of efficiency refers to the effectiveness reusability of SaaS.
- **Profitability**
  Using reusable services allow providers respond rapidly to achieve tenants' different requirements. The more satisfied tenants with providers' services, the long they will stay using provider services. Thus, there is a need to compute *Recurring Revenue* (RR), which is the annually or monthly obtained revenue from tenants. *Churn Rate* (CR), the percentage of tenants who leaving providers over a period time due to dissatisfaction with their services [65 and 75].

$$RR = \frac{Revenue\ of\ Subscribtion}{Period\ of\ Subscribtion}$$

$$Churn\ Rate\ (CR) = \frac{The\ amount\ of\ left\ tenants}{Total\ numbers\ of\ tenants * Elapsed\ Time\ Amount}$$

Unlike Recurring Revenue, decreasing churn rate affects provider profit significantly. Providing reusable SaaS services to achieve tenants' requirements aid in increasing the amount of tenants whom using provider services. The more tenants acquired by providers, the higher profitability can be gained.

- **Cost**
  It has two metrics to compute cost, which are the Reusability of Service and Reusability Saving.
  - ***Reusability of Service:***
    Service Reuse with Modification (SRM) means that service is not matching the specified requirements. Thus there is a need to calculate the cost of finding a service (CFS) in service repository, the required cost to modify the service (CMS), probability of finding a reusable service in repository (P(SF)), and the service development cost from scratch (SDC). (ERSR) is the existence of Reusable Service in Repository. As value of $P(SF)$ reduces, the number of externally used services increases.
    $$SRM = CFS + CMS + [P(SF) * SDC]$$
    $$where, P(SF) = 1 - ERSR$$
    Service Reuse without Modification (SRNM) means that service is matching the requirements. So the service modification cost will be removed to obtain the reuse cost of unmodified service and keep all the other variables as it is.
    $$SRNM = CFS + [P(SF) * SDC]$$
    The value of SRM and $SRNM$ should not exceed the service development cost from scratch.
  - ***Reusability Saving:***
    Measures the saving cost after applying reusability.
    $$Reusability\ Saving = \ CDR - \ CRR$$
    $$where, CDR\ is\ Cost\ Discard\ Reusability$$
    $$CRR\ is\ Cost\ Regard\ Reusability$$
  Decreasing a service reuse cost indicates a significant impact on the overall cost of SaaS.
- **Upgradability**
  Upgradability allows providers engaging their tenants in the provided SaaS and providing a strong application that acquires more tenants. Reusability aids providers in introducing new services quickly responding to different tenants requirements. It helps providers choosing specific services and upgrading them without affecting the other parts of the SaaS [76]. Thus, upgradability and reusability have an influence on each other. The more effective upgrades made by provider responding to tenants different requirements, the higher customer acquisition can be achieved which allows providers selling their software and increasing their revenue.
- **Response Time**
  It measures the interval time between requesting a service and receiving a service response. This measure can be used to estimate the efficiency of applying reusability for upgrading SaaS quickly responding to tenants' requirements. The shorter a provider respond to tenants with the required requirements, the more tenants will be acquired and the higher profit will be gained for provider.
- **Execution Time**
  It measures the executable time of a service. This measure can be used to estimate the efficiency of applying reusable service to SaaS. Minimal execution time permits a provider to achieve tenants requirements efficiently, gather more tenants, and increase revenue.
- **Integrity**
  The SaaS is required to have authentication measures in place at all reusable services should not interfere with each other. Moreover, any customization or configuration on SaaS must be protected against unauthorized modification.

VI. COMPARISON WITH RELATED WORKS

Most of current research works are targeted the quality attributes and metrics of object oriented [77], component oriented [4, 34] and service oriented architecture based system [7, 17, 35, 36, and 39-41]. However, a few literature studies partially addressed the quality of Software as a Service such as [19, 24, 33, 35, 42, and 44]. This section will propose a comparison with the previous works.

- The work in [24] proposed many quality attributes for SaaS for both users and providers. However, the author work excluded the business view and repeated some attributes that can be sub-categorized under a main attribute. Moreover, they did not mention the reusability as a quality attribute of SaaS. In addition, they mixed some definitions of the quality of service (QoS).
- The work in [19] had divided SaaS quality into three attributes and specified three roles for each attributes. However, the authors did not illustrate or mention the reusability as quality attribute. In addition, they did not provide a metrics to measure quality of SaaS.
- In [39] the authors measured the functional reusability of services based on their relevance. However, they only estimated the impact and applicability of services in their environment to estimate the number of compositions that may contain a service. In addition, the proposed work had handled reusability and neglected the reuse capability. Moreover, the author only explained reusability in service oriented not in SaaS.
- The authors of [17] explored and surveyed the available reusability metrics of the existed works. However, the most number of metrics had handled reusability from service-oriented architecture and component oriented views. The authors did not handle reusability in SaaS. Moreover, part of metrics had taken design characteristics of service and the other part had explored the quality characteristic of service. Thus, a further study and a proper reusability metrics are needed to measure SaaS quality. In addition, a standardized quality model for SaaS is required.
- The work in [35] had provided a quality model for SaaS. This model defines SaaS key features and derives quality attributes from these features. The authors defined metrics for the attributes. However, their quality model of SaaS as well as metrics need more investigations to extend and evaluate the quality of SaaS.
- In [33] the author provided set of quality attributes for SaaS. However, he did not include reusability in the defined attributes. In addition, he did not measure the effectiveness of the attributes.
- The work in [4] proposed and validated metrics for reusability of component based system. However, it did not target SaaS or support the provider view. In addition, the authors' reusability metrics need to be extended to obtain SaaS capabilities.
- The authors in [34] proposed metrics and attributes that influencing component based software reusability. However, their survey did not address reusability of software as a service.
- The authors of [7] proposed a quality model for evaluating reusability of SOA services. Nevertheless, their quality model should be revised to include SaaS reusability attributes. In addition, they did not handle provider side. Moreover, their proposed metrics did not properly fit suite reusability attribute.
- In [77] the authors presented a model that combines reusability and agile features to provide reusable objects. Although, the work neither considered reusability in SaaS nor in services. The authors only considered design and repository issues in reusability. Moreover, they did not mention the quality attributes of reusability.
- In [40] the authors surveyed and evaluated the quality models of web services. However, they did not handle reusability attributes of SaaS. Unlike the work in [35] which provided a quality model and metrics for evaluating SaaS. However, the business side had not been considered in the model. In addition, SaaS has many other quality attributes, which are not handled in the model.
- The work in [36] identified non-functional attributes and categorized them into multilevel stakeholders. However, the authors work only handled service oriented architecture reusability and neglected the other attributes of service such as functionality attributes. Furthermore, the business side had not been accounted in the model.
- The proposed model has several advantages. For instance, it identified SaaS common quality attributes and extracted from them the critical attributes of SaaS of reusability. The critical attributes have been divided into basic and optional to fulfill different requirements of the provider proposed applications. Moreover, it addressed some metrics to be used as a reusability measurement indicator for SaaS. In addition, the proposed reusability quality attributes and metrics not only targeting provider but they also included the business side.

VII. CONCLUSION AND FUTURE WORK

SaaS is one of cloud services that emerged as an effective reuse paradigm. In this paper, quality attributes and metrics for evaluating SaaS based on provider and business value have been proposed. First, the common quality attributes of SaaS, Services, and Software have been mentioned. Second, from the common attributes, the reusability and customizability quality attributes have been extracted. Third, from step two the critical quality attributes of reusability of SaaS have been proposed. Six critical attributes have been identified each one of them has basic and optional attributes. Fourth, the measurement for each sub-critical attribute has been specified. Finally, to show the effectiveness of the proposed critical attributes and measures a comparison with current quality models has been performed.

Currently, to depict validity of the critical quality attributes and its measures and to evaluate them, an assessment as well as weight for each metric of each attribute will be identified. Moreover, a questionnaire will be conducted to demonstrate the importance of the proposed model and metrics.

REFERENCES

[1] *The NIST Definition of Cloud Computing*, NIST 800-145, 2011.

[2] W. Lee and M. Choi. (2012, June) A Multi-tenant Web Application Framework for SaaS. Presented at IEEE 5th International Conference on Cloud Computing (CLOUD). [Online]. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6253608

[3] F. Chong and G. Carraro, (2006, April). Architecture Strategies for Catching the Long Tail. [Online]. Available: http://msdn.microsoft.com/en-us/library/aa479069.aspx

[4] K. Amit, C. Deepak and K. Avadhesh. (2014, May). Empirical Evaluation of Software Component Metrics. *The International Journal of Scientific & Engineering Research*. [Online]. *5(5)*, pp. 814-820. Available: http://www.ijser.org/researchpaper%5CEmpirical-Evaluation-of-Software-Component-Metric.pdf

[5] T. Karthikeyan and J. Geetha. (2011, Dec). A Software Standard and Metric Based Framework for Evaluating Service Oriented Architecture. *The International Journal of Engineering Research and Applications (IJERA)*. [Online]. *1(4)*, pp. 1740-174. Available: http://www.ijera.com/papers/Vol%201%20issue%204/BR01417401743.pdf

[6] O.P. Rotaru and M. Dobre. (2005, Oct). Reusability Metrics for Software Components. Presented at the 3rd ACS/IEEE International Conference on Computer Systems and Applications. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1387023

[7] S.W. Choi and S.D. Kim. (2008, July). A Quality Model for Evaluating Reusability of Services in SOA. Presented at the 10th IEEE Conference on E-Commerce Technology. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4785077

[8] S. Singh and R. Singh. (2012, Oct). Reusability Framework for Cloud Computing. *The international Journal Of Computational Engineering Research (IJCER)*. [Online]. *2(6)*, pp. 169-177. Available: http://arxiv.org/ftp/arxiv/papers/1210/1210.8011.pdf

[9] K.S. Jasmine and R. Vasantha, "A New Process Model for Reuse Based Software Development Approach," in *Proc. WSE*, London., 2008, pp. 251-254.

[10] G. Gui. (2006, Jan). Component Reusability and Cohesion Measures in Object-Oriented Systems. Presented at the ICTTA'06 2nd IEEE. [Online]. Available: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1684869&abstractAccess=no&userType=inst

[11] C. Luer, "Assessing Module Reusability," in *ACoM*., Minneapolis., Minnesota, 2007, pp. 7

[12] F. William and T. Carol. (1996, June). Software Reuse: Metrics and Models. *The Journal of ACM Computing Surveys (CSUR)*. [Online]. *28(2)*, pp. 415-435. Available: https://vtechworks.lib.vt.edu/bitstream/handle/10919/19897/TR-95-07.ps?sequence=1

[13] W.B. Frakes and K. Kang. (2005, Aug). Software Reuse Research: Status and Future. Presented at the IEEE Transaction Software Engineering. [Online]. Available: http://www.computer.org/csdl/trans/ts/2005/07/e0529-abs.html

[14] O. Sang, L. Hyun, and K. Soo. (2011, Oct). A Reusability Evaluation Suite for Cloud Services. Presented at the Eighth IEEE International Conference on e-Business Engineering. [Online]. Available: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=6104606&abstractAccess=no&userType=inst

[15] S. Sarbjeet, T. Manjit, S. Sukhvinder, and S. Gurpreet. (2010, Oct). Software Engineering - Survey of Reusability Based on Software Component. *The International Journal of Computer Applications (IJCA)*. [Online]. *8(12)*, pp. 39-42. Available: http://www.ijcaonline.org/volume8/number12/pxc3871736.pdf

[16] S.R. Schach. (2010, Jun 4). Object-Oriented and Classical Software Engineering. (8th ed.) [Online] Available: http://www.ucccs.info/ucc/ucc3/ucc_folder/cs3500/books/Object%20Oriented%20And%20Classical%20Software%20Engineering%208th%20Edition%20V413HAV.pdf

[17] T. Karthikeyan and J. Geetha. (2012, May). A Study and Critical Survey on Service Reusability Metrics. *The International Journal of Information Technology and Computer Science*. [Online]. *4(5)*, pp. 25-31. Available: http://www.mecs-press.org/ijitcs/ijitcs-v4-n5/v4n5-4.html

[18] SP Consortium, "Ada 95 quality and style guide: Guidelines for Professional Programmers," Ada95. Department of Defense Ada Joint Program Office., Herndon., VA. Virginia, Rep. SPC-94093-CMC, Oct. 1995.

[19] B. Sarbojit and J. Shivam. (2014, Jan). A Survey on Software as a Service (SaaS) using quality model in cloud computing. *The International Journal Of Engineering And Computer Science (IJECS)*. [Online]. *3(1)*, pp. 3598-3602. Available: http://www.techrepublic.com/resource-library/whitepapers/a-survey-on-software-as-a-service-saas-using-quality-model-in-cloud-computing/

[20] S. Arun, K. Rajesh and P. S. Grover. (2007, Jan). A Critical Survey of Reusability Aspects for Component-Based Systems. *The International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*. [Online]. *1(9)*, pp. 420-424. Available: http://waset.org/publications/3086/a-critical-survey-of-reusability-aspects-for-component-based-systems

[21] Loosely Coupled. "*Define Meaning Of Customization,"* looselycoupled.com. [Online]. Available: http://looselycoupled.com/glossary/customization. [Accessed Feb. 12, 2016].

[22] C. Lizhen, W. Haiyang, J. Lin, and H. Pu. (2010, Dec). Customization Modeling based on Metagraph for Multi-Tenant Applications. Presented at the 5th International Conference on Pervasive Computing and Applications (ICPCA). [Online]. Available: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5704108&abstractAccess=no&userType=inst

[23] H. Mahdavi, Sara, G. Matthias and A. Paris. (2013, Feb). Variability in quality attributes of service-based software systems: A systematic literature review. Presented at in the Conference of Information and Software Technology. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0950584912001772

[24] K. Khanjani, N. Wan, A. AbdulAzim, and Md. AbuBakar. (2014, Oct). SaaS Quality of Service Attributes. *The Journal of Applied Sciences*. [Online]. *14(24)*, pp. 3613-3619. Available: http://scialert.net/qredirect.php?doi=jas.2014.3613.3619&linkid=pdf

[25] F. Stefan Frey, L. Claudia, and R. Christoph. (2013, Oct). Key Performance Indicators for Cloud Computing SLAs. Presented at the Fifth International Conference on Emerging Network Intelligence (EMERGING). [Online]. Available: http://www.chinacloud.cn/upload/2013-10/13101108464452.pdf

[26] Q. Meng, Y. Ying, and W. Chen. (2013, July). Systematic Analysis of Public Cloud Service Level Agreements and Related Business Values. Presented at the 10th IEEE International Conference on Services Computing. [Online]. Available: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=6649763&abstractAccess=no&userType=inst

[27] P. Berander, L.O. Damm, J. Eriksson, T. Gorschek, K. Henningsson, P. Jönsson and P. Tomaszewski. (2005, June). *Software quality attributes and trade-offs*. (1st ed). [Online]. Available: https://www.uio.no/studier/emner/matnat/ifi/INF5180/v10/undervisningsmateriale/reading-materials/p10/Software_quality_attributes.pdf

[28] Selecting a cloud provider, CSMIC FramewokV2.1, 2014.

[29] Felix and K. Mark, "Understanding Quality Attributes". in *Software Architecture in Practice*, 3th ed. Westford: Addison-Wesley, 2013, Ch. 4, Sec. 4.3, pp. 61-77

[30] S. Chriss (2011, March. 14). "*Cloud Services Measurement Consortium Gains Critical Membership*". CMU [Online]. Available: https://www.cmu.edu/news/archive/2011/March/march14_cloudconsortium.shtml. [Accessed: Feb. 13, 2016].

[31] ISO 9126: The Standard of Reference, ISO/IEC 9126, 1991.

[32] ISO 25000 Standards, ISO/IEC 25010, 2015.

[33] B. Lukas. (2013, Jan). Quality of Service Attributes for Software as a Service. *The Journal of Systems Integration*. [Online]. *4(3)*, pp. 38-47. Available: http://www.si-journal.org/index.php/JSI/article/viewFile/166/126

[34] M. Marko and S. Zlatko. (2015, Oct). Reusability Metrics of Software Components: Survey. Presented at the Central European Conference on Information and Intelligent Systems (CECIIS). [Online]. Available: https://bib.irb.hr/datoteka/782030.Reusability_metrics_of_Software_Components.pdf

[35] L. Jae, L. Jung, C. Du and K. Soo. (2009, Dec). A Quality Model for Evaluating Software as a Service in Cloud Computing. Presented at the Seventh International Conference on Software Engineering Research, Management and Applications (ACIS). [Online]. Available: http://dl.acm.org/citation.cfm?id=1730243

[36] G. Shanmugasundaram, V. Prasanna, and D. Punitha. (2012, Jul). A Comprehensive Model to achieve Service Reusability for Multilevel Stakeholders using Non-Functional attributes of Service Oriented Architecture. Presented at the International Conference on Intelligence Computing (ICONIC'12). [Online]. Available: http://arxiv.org/abs/1207.1173

[37] S. Neha, D. Surender, and A.S. Manjot. (2013, Jul). An Empirical Review on Factors Affecting Reusability of Programs in Software Engineering. *The International Journal of Computing and Corporate Research (IJCCR)*. [Online]. *3(4)*, pp. 7. Available: http://www.ijccr.com/July2013/5.pdf

[38] A. Ashraf, "Quality Attributes Investigation for SaaS in Service Level Agreement," M.S. thesis, Dept. Mathematics and Computer Science., Çankaya Univ., Turkey, Ankara, 2015.

[39] M. Fleix. (2015, Jan). A Metric for Functional Reusability of Services. Presented at the 14th International Conference on Software Reuse (ICSR). [Online]. Available: http://link.springer.com/chapter/10.1007%2F978-3-319-14130-5_21

[40] O. Marc, M. Jordi and F. Xavier. (2014, Oct). Quality models for web services: A systematic mapping. *The Journal of Information and Software Technology*. [Online]. *56(10)*, pp. 1167-1182. Available: http://kczxsp.hnu.edu.cn/upload/20141022111314350.pdf

[41] R. Sindhgatta, B. Sengupta, and K. Ponnalagu. (2009, Nov). Measuring the Quality of Service Oriented Design," Presented at the 7th International Joint Conference on Service-Oriented Computing (ICSOC-ServiceWave'09). [Online]. Available: http://link.springer.com/chapter/10.1007%2F978-3-642-10383-4_36

[42] F. Nemésio, P. Clarindo, B. Paulo, Z. Andre, and B. Urlan. (2013, June). SaaSQuality-A Method for Quality Evaluation of Software as a Service (SaaS). *The International Journal of Computer Science & Information Technology (IJCSIT)*. [Online]. *5(3)*, pp. 101-117. Available: http://www.academia.edu/3835761/SaaSQuality_-A_method_for_Quality_Evaluation_of_Software_as_a_Service_SaaS_

[43] A. Ibraheem, A. Abdallah, and Y. Mohd. (2014, Jan). Taxonomy, Definition, Approaches, Benefits, Reusability Levels, Factors and Adaption of Software Reusability: A Review of the Research Literature. *Journal of Applied Sciences*. [Online]. *14(20)*, pp. 2396-2421. Available: http://www.scialert.net/fulltext/?doi=jas.2014.2396.2421

[44] L. Hyun, and K. Soo. (2009, Dec). A Systematic Process for Developing High Quality SaaS Cloud Services. Presented at the International Conference of Cloud Computing. [Online]. Available: http://link.springer.com/chapter/10.1007%2F978-3-642-10665-1_25

[45] H. Jeong, Y. Hwa, and H. Bong, "The Identification of Quality Attributes for SaaS in Cloud Computing," *AMM. Applied Mechanics and Materials*, vol. 300-301, no. 1, pp. 689-692, Feb. 2013.

[46] W.N.T. Alwis, and C.D. Gamage. (2014, April). A Multi-tenancy Aware Architectural Framework for SaaS Application Development. *Journal of the Institution of Engineers*. [Online]. *46(3)*, pp. 21-31. Available: http://engineer.sljol.info/article/download/6782/5280/

[47] H. Washizaki, H. Yamamoto, and Y. Fukazawa, "A Metrics Suite for Measuring Reusability of Software Components," in Proc. *9th ISMS*, 2003, pp. 211-223.

[48] A. Sharma, PS. Grover, R. Kumar. (2009, March). Reusability Assessment for Software Components. *ACM SIGSOFT Software Engineering*. [Online]. *34(2)*, pp. 1-6. Available: http://sdiwc.us/digitlib/journal_paper.php?paper=00000045.pdf

[49] S. Shrddha, N. W. Nerurkar, and A. Sharma. (2010, July). A Soft Computing based Approach to Estimate Reusability of Software Components. *ACM SIGSOFT Software Engineering*. [Online]. *35(4)*, pp. 1-5. Available: http://gmm.fsksm.utm.my/~mrazak/articles/ACM--Sagar_2010--A_Soft_Computing_Based_Approach_to_Estimate_Reusability_of_Software_Components.pdf

[50] J. Aman, and G. Deepti, "Estimation of Component Reusability by Identifying Quality Attributes of Component: a Fuzzy Approach," in *Proc. CCSEIT '12*, 2012, pp. 738-742.

[51] T.R. Stefan, and A. Urs, "Applying Software Product Lines to Create Customizable Software-as-a-Service Applications," in *Proc. ISPLC*, 2011, pp. 1-4

[52] E.S. Cho, M.S. Kim, and S.D. Kim, "Component Metrics to Measure Component Quality," in *Proc. 8th APSEC*, 2001, pp. 419 – 426.

[53] SD. Kim, JH. Park. (2003, Feb). C-QM: A Practical Quality Model for Evaluating COTS Components. Presented at the international Multi-Conference on Applied Informatics. [Online]. Available: https://www.researchgate.net/publication/221277611_C-QM_A_Practical_Quality_Model_for_Evaluating_COTS_Components

[54] L. O'Brien, P. Merson, and L. Bass, "Quality Attributes for Service-Oriented Architectures," in *Proc. SDSOA*, 2007, pp. 1-38.

[55] B. Behkamal, M. Kahani, and M. K. Akbari. (2009, Mar). Customizing ISO 9126 Quality Model for Evaluation of B2B Applications. *Information and software technology*. [Online]. *51(3)*, pp. 599-609. Available: http://profdoc.um.ac.ir/pubs_files/p11004966.pdf

[56] C. Janz (2015, April. 24). "*Key Revenue Metrics for SaaS Companies*". Christoph Janz [Online]. Available: http://christophjanz.blogspot.com.eg/2015/04/key-revenue-metrics-for-saas-companies.html. [Accessed: Feb. 13, 2016].

[57] V. Naganathan, and S. Sankarayya, (2012, March). Overcoming Challenges associated with SaaS testing. *Infosys, Building Tomorrow's Enterprise*. [Online]. Available: https://www.infosys.com/IT-services/independent-validation-testing-services/white-papers/Documents/overcoming-challenges-saas-testing.pdf

[58] X. Chen, and P. Sorenson, (2011, March). Quality-Based Co-Value in SaaS Business Relationships. *Technology Innovation Management Review*. [Online]. Available: http://timreview.ca/article/428

[59] D. Skok (2013, Feb. 10). "*SaaS Metrics 2.0 – A Guide to Measuring and Improving what Matters*". For Entrepreneurs [Online]. Available: http://www.forentrepreneurs.com/saas-metrics-2/. [Accessed: Feb. 13, 2016].

[60] A. Chang (2014, Sept. 17). "*7 Critical Cloud Service Attributes*". Network Computing [Online]. Available: http://www.networkcomputing.com/cloud-infrastructure/7-critical-cloud-service-attributes/1852724555. [Accessed: Feb. 13, 2016].

[61] X. Chen, A. Srivastava, and P. Sorenson, "Towards a QoS-Focused SaaS Evaluation Model," in *Cloud Computing and Software Services Theory and Techniques*, 1st ed., vol. 13, A. Syed and I. Mohammad, Ed. Alberta: CRC Press, 2011, Ch. 16, sec. 16.4, pp. 389–407.

[62] M. Qiu, Y. Zhou, and C. Wang. (2013, July). Systematic Analysis of Public Cloud Service Level Agreements and Related Business Values. Presented at the IEEE International Conference on Services Computing (SCC). [Online]. Available: https://sro.library.usyd.edu.au/handle/10765/102011

[63] Corporate Management. "*Software as a Service (SaaS)*," itinfo.am. [Online]. Available: http://www.itinfo.am/eng/software-as-a-service/. [Accessed: Feb. 13, 2016].

[64] G. McKeown (2014, Dec. 12). "*SaaS vs. On-Premise – what suits your organisation?,*". Gael Quality [Online]. Available: http://www.gaelquality.com/customer-support-blog/december-2014/saas-vs-on-premise-–-what-suits-your-organisation. [Accessed: Jan. 2, 2016].

[65] D. Skok (2010, Feb. 10). "*SaaS Metrics – A Guide to Measuring and Improving What Matters*". For Entrepreneurs [Online]. Available: http//www.forentrepreneurs.com/saas-metrics/. [Accessed: Feb. 13, 2016].

[66] S. K. Garg, S. Versteeg, and R. Buyya. (2013, June). A Framework for Ranking of Cloud Computing Services. *Future Generation Computer Systems*. [Online]. *29(4)*, pp. 1012-1023. Available: http://www.cloudbus.org/papers/RankingClouds-FGCS.pdf

[67] G. Khatleen (2009, June. 15). "*Building your Business Case for Best Practice IT Service Delivery*". Oracle On Demand [Online]. Available: http://www.oracle.com/us/products/ondemand/collateral/outsourcing-center-business-case-344142.pdf. [Accessed: Feb. 13, 2016].

[68] P. Anderson (2015, June. 25). "*The Six Attributes of Cloud based Applications that Explain Businesses' High Inclination Towards Them*". Customer Think [Online]. Available: http://customerthink.com/the-six-attributes-of-cloud-based-applications-that-explain-businesses-high-inclination-towards-them/. [Accessed: Feb. 13, 2016].

[69] Open Group. "*Cloud Computing for Business: Why Cloud?,*" opengroup.org. [Online]. Available: http://www.opengroup.org/cloud/cloud/cloud_for_business/why.htm. [Accessed: Feb. 13, 2016].

[70] T. Huynh, Q. Pham, and V. Tran, "The Reusability and Coupling Metrics for Service Oriented Softwares," Unpublished.

[71] A. Khoshkbarforoushha, P. Jamshidi, F. Shams. (2010, May). A Metric for Composite Service Reusability Analysis. WETSoM'10. [Online]. Available: http://www.computing.dcu.ie/~pjamshidi/PDF/ICSE-WETSoM2010.pdf

[72] N. Mohd, A. K. Raees, and M. Khurram. (2010, April). A Metrics Based Model for Understandability Quantification. *Journal of Computing*. [Online]. *2(4)*, pp. 90-94. Available: http://arxiv.org/ftp/arxiv/papers/1004/1004.4463.pdf

[73] A. Kazemi, A. Rostampour, A.N. Azizkandi, and H. Haghighi. (2011, June). A Metric Suite for Measuring Service Modularity. Presented at International Symposium on Computer Science and Software Engineering (CSSE). [Online]. Available: https://www.researchgate.net/publication/252022502_A_metric_suite_for_measuring_service_modularity

[74] A. Jugoslav and T. Vladimir. (2012, Sept). Metrics for Service Availability and Service Reliability in Service-oriented Intelligence Information System. Presented at the International Conference ICT Innovations. [Online]. Available: http://eprints.ugd.edu.mk/6535/

[75] Y. Joel. (2010, Sept). SaaS Metrics Guide to SaaS Financial Performance. *Chaotic-flow*. [Online]. Available e-mail: JOELYORK@CHAOTIC-FLOW.COM

[76] G. Paul (2016). "*SaaS Design Principles-Use Modular Design to Enable Upsell and Upgradability*". Catalyst Resources [Online]. Available: http://catalystresources.com/saas-blog/saas_design_principle_use_modular_design_to_enable_upsell_and_upgradability. [Accessed: Feb. 12, 2016].

[77] S. Sukhpal and C. Inderveer. (2012, July). Enabling Reusability in Agile Software Development. *The International Journal of Computer Applications*. [Online]. *50(13)*, pp. 33-40. Available: http://arxiv.org/ftp/arxiv/papers/1210/1210.2506.pdf

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 3, March 2016

A Model Driven Regression Testing Pattern for Enhancing Agile Release Management

Maryam Nooraei Abadeh,
Department of Computer Science, Science and Research Branch, Islamic Azad University, Tehran, Iran

*Seyed-Hassan Mirian-Hosseinabadi
Department of Computer Science, Sharif University of Technology, Tehran, Iran

***Abstract*- Evolutionary software development disciplines, such as Agile Development (AD), are test-centered, and their application in model-based frameworks requires model support for test development. These tests must be applied against changes during software evolution. Traditionally regression testing exposes the scalability problem, not only in terms of the size of test suites, but also in terms of complexity of the formulating modifications and keeping the fault detection after system evolution. Model Driven Development (MDD) has promised to reduce the complexity of software maintenance activities using the traceable change management and automatic change propagation. In this paper, we propose a formal framework in the context of agile/lightweight MDD to define generic test models, which can be automatically transformed into executable tests for particular testing template models using incremental model transformations. It encourages a rapid and flexible response to change for agile testing foundation. We also introduce on-the-fly agile testing metrics which examine the adequacy of the changed requirement coverage using a new measurable coverage pattern. The Z notation is used for the formal definition of the framework. Finally, to evaluate different aspects of the proposed framework an analysis plan is provided using two experimental case studies.**

Keywords *Agile development. . Model Driven testing. On-the fly Regression Testing. Model Transformation. Test Case Selection.*

I. INTRODUCTION

The Model Driven Architecture (MDA) paradigm enhance traditional development discipline with defining a platform independent model (PIM), which is followed by manually or automatically transforming it to one or more platform specific model (PSM), and completed with a code generation from PSMs [1]. The MDA profits, e.g., abstraction modeling, automatic code generation, reusability, effort reduction and efficient complexity management can be influenced to all phases of the software lifecycle. To get all advantages of MDD, it is essential to use it in an agile way, involving short iterations of development with enough flexibility and automation. Because it is so easy to add functionality when using MDD, you will not be the first one ending up with a ‘concrete-model’. On the other hand Model transformation and traceability, as two key concepts MDA, provide an automatic maintenance management’s ability in a more agile and rapid release environment MDD to make it possible to show the results of a model change almost directly on the working application. Agile MDA principles, e.g., alliance testing, immediate execution, racing down the chain from analysis to implementation in short cycles should be applied in short incremental, iterative cycles. To support agile changes and at different levels of abstraction, e.g., requirement specification, design, implementation using manual or semi-automated refactoring approaches, efficient change management supports induced changes using update propagation. Update propagation has been essentially used to provide techniques for efficient traceable and incremental view maintenance and integrity checking in different phases of software development. Incremental refactoring of model improves the development and test structure to early defect fault introduced through evolution more precise.

---

*Author for correspondence

When developing safety-critical software systems there is, however, a requirement to show that the set of test cases, covers the changes to enable the verification of design models against their specifications. Automatic update propagation can enhance regression testing in a formal way to reduce the associated efforts. The purpose of regression testing is to verify new versions of a system to prevent functionality inconsistencies among different versions. To avoid rerunning the whole test suite on the updated system various regression test selection techniques have been proposed to exercise new functionalities according to a modified version [2]. Propagating design changes to the corresponding testing artifacts leads to a consistent regression test suite. Normally this is done by transforming the design model to code, which is compiled and executed to collect the data used for structural coverage analysis. If the structural code coverage criteria are not met at the PSM level, additional test cases should be created at the PIM level.

The proposed approach is a MDT version of agile development. The motivation with the proposed approach is that instead of creating extensive models before writing source code you instead create agile models which are just barely good enough that drive your overall development efforts. Agile model driven regression testing is a critical strategy for scaling agile software development beyond the small changes during the stages of agile adoption. It provides continuous integration, maintenance and testing even with platform changes.

As a technical contribution of the current paper, we use the Z specification language [3] not only for its capability in software system modeling, development and verification, but also for its adaptation in formalizing MDD concepts, e.g., model refactoring, transformation rules, meta-model definition and refinement theory to produce a concrete specification. Besides, verification tools such as CZT [4] and Z/EVES [5] have been well-developed for type-checking, proofing and analyzing the specifications in the Z-notation. Although OCL can be used to answer some analysis issues in MDD, it is only a specification language and the mentioned mechanisms for consistency checking are not supported by OCL.

Finally, a main challenge which is investigated in this paper is: how will abstract models be tested in an agile mythology? How can use MDA-based models to handle the inherent complexities of legacy system testing? Is developing these complex models really more productive than other options, such as agile development techniques?

The rest of the paper is organized as follows: Section 2 reconsiders the related concepts. Section 3 extends the formalism for platform independent testing. In Section 4 the agile regression testing is introduced. Section 5 introduces on-fly agile (regression) testing framework. The practical discussion and analysis of the framework are provided in Section 6. Section 7 reviews the related works and compares the similar approaches to our work. Finally, Section 8 concludes the paper and gives suggestions for future works.

II. Preliminaries

In this section, we review some preliminary concepts that are prerequisites for our formal framework.

A. *Regression test selection, minimization and prioritization*

Regression testing as a testing activity during the system evolution and maintenance phase can prevent the contrary effects of the changes at different levels of abstraction. Important issues have been studied in regression testing to keep and maximize the value of the accrued test suite are test case selection, minimization and prioritization. Regression Test Selection Techniques (RTSTs) select a cost-effective subset of valid test cases from previously validated version to exercise the modified parts of a model/program. A RTST essentially consists of two major activities: identifying affected parts of a system after the maintenance phase and selecting a subset of test cases from the initial test suite to effectively test the affected parts of the system. A suitable coverage by a number of test cases is needed that detects new potential faults. A well-known classification of regression test cases is suggested in [2] which classifies test suites into obsolete, reusable and retestable test cases. Obsolete test cases are invalid for the new version and should be removed from the original test pool and two others are still valid to be rerun. Test case selection, or the regression test selection problem, is essentially similar to the test suite minimization problem; both problems are about choosing a subset of test cases from the test suite. The key difference between these two approaches in the literature is whether the focus is upon the changes in the system under test. Test suite minimization is often based on metrics such as coverage measured from a single version of the program under test. By contrast, in regression test selection, test cases are selected because their accomplishment is relevant to the changes between the previous and the current version of the system under test. Minimization techniques aim to reduce the size of a test suite by eliminating redundant test cases. Effective minimization techniques keep coverage of reduced subset equivalent as the original test suite while reducing the maintenance costs and time. Compared to test case selection techniques that also attempt to reduce the size of a test suite, the selection is not only focuses on the current version of a system, but the most of selection techniques are change-aware [6].

Finally, test case prioritization techniques attempt to schedule test cases in such an order that meet desired properties, such as fault detection, at an earlier stage. The important issues of regression testing at the platform independent level will be solved by systematic analyzing of system specifications and enhanced by following test strategies, e.g., coverage criteria to adequately cover demanded features of the updated models. A main metric that is often used as the prioritization criterion in coverage-based prioritization is the structural coverage. The intuition behind the idea is that early maximization of the structural coverage will also increase the ability of early maximization of the fault detection [7].

III. Agile Testing in the Model Driven Context

Agile methodologies have permanently changed the traditional way of software development by focusing on changes via accepting the idea that requirements will evolve throughout a software development. It aims to organize adaptive planning, evolutionary development, complex multi-participant software development while achieving fast delivery of quality software, better meeting customer requirements and rapid and flexible response to changes. The emphasis on testing approaches has been rising with the widespread application of agile methodologies [8]. Iterative testing, especially Test-Driven Development (TDD) [9], can be a foundation for quality assurance in all methods. Through early integrating testing into the main development lifecycle and focusing on automated testing, agile/lightweight methodologies aim to achieve a low defect. Besides detecting faults, it is expected that the development and maintenance phases will be flexible enough to dynamically react on retesting changing requirements [10].

The utilization of models, especially the use of practical, object-oriented models (e.g., UML) in agile processes can enhance the opportunity of benefiting from MBT in agile/lightweight approaches. Unfortunately, most of MBT research does not provide general, direct solutions to problems in agile testing. We have identified a major area of agile/lightweight testing in which MBT may lead to incremental agile testing solutions. It leads to achieve a higher level of abstraction is often promised as one of the major benefits of MDT. Since test cases are usually developed in an ad hoc and incremental manner in agile/lightweight methodologies, the benefits of MDD can be significant in this context. It deals with how MDT can provide overall direction to the testing process and provide meaningful test coverage goals.

Using model transformations and traceability links, we can bridge the gap between MDD and Model Driven Testing (MDT). MDT promotes MDA advantages to software testing. MDT is started at a high level of abstraction to derive abstract test cases, then transforms the abstract test cases to concrete test code using stepwise refinement. It can reduce maintenance and debugging problems by providing traceability links in a forward and backward direction simultaneously. The general solution for agile model driven testing may be use the meta-model based traceability/transformation between a system design model and its test requirement model. The source and target instance models conform to their corresponding meta-models. Refinement rules are used to automatic movement to concrete environment step by step, e.g., to enrich a PIT to PST required test specific properties must be added. When a test case detects any type of ‘mistake’ at different levels of abstraction, it is possible to follow traceability links to identify and resolve its origins at an early stage of the software development lifecycle that results in enhancing the software quality and reducing the maintenance cost. In this paper, we extend MDT philosophy to expose model driven regression testing. Promoting the philosophy of MDA to the maintenance phase can reinforce worthy domains like “Model Driven Software Maintenance” which leads to cost-effective configuration management. Therefore, we want to raise the level of abstraction for the logical coverage analysis to be performed at the same level as the design model verification, i.e., the PIM level.

An important type of MT uses incremental graph pattern matching to synchronize model, called incremental MT. It observes changes to the source model, and then propagates those parts of the source that changed to the target model. This synchronization process does not re-generate the whole artifacts, but only updates the models according to the changes while batch transformation discards transformation results. Thus, if the transformation is defined between design and test meta-models, the original model changes will be propagated to test model to update the regression test suite. Semantically, it creates test cases in the target if such test cases do not exist, it modifies test cases if such test cases exist, but have changed, and it deletes test cases which traverse inaccessible model elements. Different kinds of model transformations may be useful during model evolution, but it is necessary to be complemented by the necessary efforts of inconsistency management, to deal with possible inconsistencies that may arise in a model after its transformation. When the consistency problem and model transformations are integrated, different interesting transformations can be defined [11]. In our approach, bidirectional consistency-preserving transformations, described by Stevens [12], are desirable. Bidirectional and change propagating model transformations are two important types of transformations which implicitly try to keep the consistency and symmetry between source and target models.

Continuously verifying changes at the abstract level and fixing problems as they happen in development leads to shorter throughput times in load and performance tests because a higher-quality software product is already tested.

In addition, traceable change management and automation in regression testing make it possible to automate a large number of testing tasks, which can then be regularly implemented by an automatic refinement process.

IV. BRIEFLY ON THE Z-NOTATIONS

Z [3] is a formal specification language based on the standard mathematical foundation with an effective structuring mechanism for precisely modeling, specifying and analyzing computing systems. Some formal languages, e.g., Z, provide a particular template to specify systems. The main ingredient in Z is a way of decomposing a specification into small pieces called schemas. The schemas can be used to explain structured details of a specification and can describe a transformation from one state of a system to another. Also, by constructing a sequence of specifications, each containing more details than the last, a concrete specification that satisfies the abstract ones can be reached [3]. In more realistic projects, such a separation of concerns is essential to decrease the complexity. Later, using the schema language, it is possible to describe different aspects of a system separately, then relate and combine them.

The symbols used in the Z-notation is principally similar to common mathematical symbols, e.g., notation for number set ($\mathbb{N}, \mathbb{Z}$ and $\mathbb{R}$), set operations ($\cup, \cap, \times$ and \), quantifiers ($\forall, \exists, \nexists$ and $\exists_1$), first order logic ($\wedge, \vee$ and $\Longrightarrow$). The power set of A is shown by $\mathbb{P}$ A. A relation R over two sets X and Y is a subset of the Cartesian product, formally, R $\in$ (X $\times$ Y). The domain and the range of R are denoted by **dom** R and **ran** R respectively. The domain and range restriction of a relation R by a set P are the relation obtained by considering only the pairs of R where respectively the first elements and the second elements are members of P, formally P $\lhd$ R and R $\rhd$ P. Z supports various mathematical functions, e.g., total, injective, subjective, and bijective. In addition to the sets and logic notations, Z presents a schema notation as an organized pattern for structuring and encapsulating pieces of information in two parts: declaration of variables and a predicate over their values. In more realistic projects, such a separation of concerns is essential to decrease the complexity. Using the schema language, it is possible to describe different aspects of a system separately, then relate and combine them. The schemas can be used to explain structured details of a specification and can describe a transformation from one state of a system to another, with related input and output variables. State variables before applying the operation are termed pre-state variables and are shown undecorated. The corresponding post-state variables have the same identifiers with a prime (‘) appended. The name of input and output variables should be ended by a query (?) and exclamation (!) respectively. In a Z operation schema, a state variable is implicitly equated to its primed post-state unless included in a Δ-list. A Δ-list is a schema including both before and after change variables. Z specifications are validated using CZT. The essential notations of the Z specification language are mentioned in Table 1. Supporting different abilities in Z/EVES include: type checking: syntax of object being specified, precise object definition: underlying syntax conformance, invariant definition: properties of modeled object and pre/post conditions: semantics of operations.

TABLE I
SOME ESSENTIAL NOTATIONS OF Z SPECIFICATION LANGUAGE

Notation	**Definition**
first e	first(e1,e2) = e1
second e	second(e1,e2) = e2
$\mathbb{P}$ e	{i : $\mathbb{W}$ | i $\subseteq$ e}

$e1 \lhd e2$	$\{i : e2 \mid first\ i \in e1\}$
$e1 \ntriangleleft e2$	$\{i : e2 \mid first\ i \notin e1\}$
$e1 \fatsemi e2$	$\{i1 : e1; i2 : e2 \mid second\ i1 = first\ i2 \bullet first\ i1 \mapsto second\ i2\}$
$e1 \oplus e2$	$((dom\ e2) \ntriangleleft e1) \cup e$
$e1 \rightarrowtail e2$	$\{i1 : P\ (e1 \times e2) \mid \forall\ i2, i3 : i1 \mid first\ i2 = first\ i3 \bullet second\ i2 = second\ i3\}$
$e1 \rightarrow e2$	$\{i : e1 \rightarrowtail e2 \mid dom\ i = e1\}$
ΔS	Shortcut for $S \wedge S'$

V. Formal overview of the approach

In this section, we propose a precise definition of behavioral model, as a typed attributed model, delta model and model refactoring which allows capturing the nature of the change unambiguously. Also, we define the dependencies between delta records in order to prepare an optimized delta model for regression testing. Finally, the abstract testing terminology based on the formalized behavioral model is described.

B. Meta-model independent behavioral model

In the MOF meta-modeling environment, each model conforms to its meta-model. We introduce a meta-model independent model integrated into the theory of labeled and typed attributed graphs [13] as the behavioral model of the testing framework. This model may conform to various meta-models, and able to apply to multiple domain-specific modeling languages. The behavioral models provide a richer semantic to represent model states, abstract test cases, coverage criteria and their relations. Also, using an independent meta-model approach can be considered as a possible candidate technique to perform a meta-model independent specification for model differences.

Definition 1 (Behavioral Model Syntax). A behavioral model is defined by a schema, named *TestModel*, where its elements are finite sets of given set *TestMMElem*. This definition behaves as a generic meta-model for defining the elements of a test model, so, each behavioral model is an instance model that its elements conform to the meta-models elements. The elements of a test model are named by their *IDs*. To take the advantages of the available theories about typed attributed model modeling in order to change propagation, type *typeID* and some attributes attributeset are assigned to each element of a behavioral model. *TestMMElem* changes are tagged by *ChgTag* to investigate coverage of different elements that are changed by update operations.

[typeID, attrID, Value, TestMMElem]

ChgTag ::= *New* | *Update* | *Delete*

DeltaOp ::= *AddMT* | *DelMT* | *UpdateMT*

attributeset == *attrID* ⤀ *Value*

TestModel

TotalElem: ℙ *TestMMElem*
type: TestMMElem ⇸ *typeID*
value: TestMMElem ⇸ *attributeset*
TagFunc: TestMMElem ⇸ *ChgTag*

C. Model Evolution

Structurally, a model of changes can be presented by two main techniques: directed deltas and symmetric deltas [14] which are different in change representing. The direct delta organizes changes on a model as a sequence of delta operations, while the symmetric delta represents the consequent of the delta operation as the set difference between two compared versions. In this paper, the directed delta technique is used to express ongoing changes on a behavioral model.

Definition 2 (Delta Operation). A delta operation is a direct model transformation for refactoring models in the same abstract syntax. Delta operations are generally divided into two main types: "Add" and "Delete". Another type of delta operations, the operation "Update", is considered as sequences of operations "Add" and "Delete" on the same element. Each delta operation has pre-and post-condition to verify a model structure before and after its execution. In our framework, post-state elements of each delta operation have the same identifiers with a prime (') appended. Graph matching is used to recognize pre- and post-condition patterns for each delta operation. Delta operations may be executed in batch, incremental or change driven manner [15]. In this paper, batch and incremental modes are desirable. The former performs all delta operations which their pre-conditions are satisfied; the latter executes delta operations one by one in an incremental mode. In the two manners, transformations update a behavioral model without the costly re-evaluation of unchanged elements. In a batch implementation, the pre-condition is the union of delta operation pre-conditions and the resulting model is the union of conflict-free post-conditions. Formally, for delta operations MTOpi that $i \geq 1$, DeltaPreCon= $\bigcup_{i=1}^{n} \text{Pre–condition}\,(\text{MTOp}_i)$ and DeltaPostCon= $\bigcup_{i=1}^{n} \text{Post–condition}\,(\text{MTOp}_i)$.

For example, consider two delta operations "AddMT" and "DelMT" on the label set of the behavioral model TestModel. To add an input label e? to the set of states of the behavioral model, the schema AddLabel is used. Its pre-conditions force that the new label e? should not exist in TestModel and the post-condition adds the label e? to the set of labels. The label-removal DelLabel deletes an existing label e? from the set of labels and restrict the domain of the functions value and type.

AddElem

ΔTestModel
s?: TestMMElem
v?: Value
attribute?: attrID

$s? \notin TotalElem$
$TotalElem' = TotalElem \cup \{s?\}$
$value' = value \cup \{(s? \mapsto \{(attribute? \mapsto v?)\})\}$
$TagFunc' = TagFunc \cup \{(s? \mapsto New)\}$

DelElem

ΔTestModel
s?: TestMMElem
type?: typeID
attribute?: attrID

$s? \in TotalElem$
$TotalElem' = TotalElem \setminus \{s?\}$

An updating model transformation can be directly defined by a schema like *UpdateElem*, or by composing of two removal and additional operation schemas, *UpdateElem* $==$*DelElem* $\fatsemi$ *AddElem*. The schema *UpdateElem* implies that for an updated label e*?*, its type and attribute set should be updated and a tag *Update* should be attached to distinguish the undergoing change on the element. So, refactoring at this level of abstraction is defined as *TestModelRefac* $==$ *ChangeElem* where *ChangeElem* $==$ *AddElem* $\vee$*DelElem* $\vee$ *UpdateElem* and its pre-condition is defined as: *DeltaPreCon* $==$ pre *ChangeElem* or *DeltaPreCon* $==$ pre *TestModelRefac*

Definition 3 (Delta Signature). Each delta signature is a 4-tuple; a changed model element, a delta operation and before and after values of the changed element. It is clear that before and after values for added and deleted elements are respectively empty. The delta signatures for our case study are defined in Section 6. Formally:

$$DeltaReq \ == TestMMElem \times DeltaOp \times Value \times Value$$

In order to simplify, all delta operations are categorized by the free type *DeltaOp* as distinct sets of direct model transformations.

$DeltaOp ::= \ AddMT \mid \ DelMT \mid \ UpdateMT$

Definition 4 (Delta Model- Abstract Syntax). Syntactically, a delta model is defined by a set of 3-tuples (*DeltaReq* $\times$*dependency* $\times$*DeltaReq*) where *DeltaReq* is a delta signature and *dependency* is an established dependency between two delta signatures. The relations between delta signatures are defined by a free type with four kinds of dependencies. . The direct dependency, called also definition/use dependency, implies that a delta signature sig1 accesses the same element which added by a delta signature sig2. A delta signature sig1 indirectly depends on delta signature sig2, if the subject element of sig1 is in a “related to”, e.g., “belongs to” or “contained/container” relation with the subject element of sig2. Finally, independent delta signatures are delta signatures which no limitation is applied on their execution order. An independent delta signature can be applied in parallel while the result of the parallel execution of delta operations is equivalent to a serial execution.

$dependency ::= \ Direct \mid \ Indirect \mid \ Indep$

It is worth noting that (as it is expressed in *Delta* predicate) *dependency* relations between *DeltaReqs* are irreflexive, asymmetric and transitive. In other words, for delta signatures *x*, *y* and *z* and dependency relations *a*: (1) *not x a x*; (2) if *x a y* implies *not y a x* and (3) if *x a y* and *y a z* implies *x a z*. Based on the mentioned descriptions, a delta model is defined by the schema *Delta*. The schema predicate also implies that deleted and updated model elements should be the members of model elements *TotalElem* while added model elements should not be the members of *TotalElem*. All delta model elements are accessible from the set *AllSig*.

We propose to optimize a delta model before propagating it to the testing phase. It leads to improve the efficiency of RTST because only test cases which traverse the optimized delta model are investigated. For example, if an element is added by a delta operation and deleted by another, no test requirement will be added to cover these changes. To manage indirect dependencies, all changes on a container should be performed earlier than changes on the corresponding contained. For an example, occurrence a couple of delta operations AddMT with direct dependency is reported as a conflict for the second additional operation, and for indirectly dependent delta signatures is managed by performing the container at the first; the same problem can happen with the tuple

(UpdateMT, AddMT) whenever impose the same model element sequentially. To manage such conflicts, we should remove the unsafe operations (the operation with an unsafe access) form a delta model.

Delta

$TestModel$
$AllSig: \mathbb{P}\ DeltaReq$
$SIGSet: \mathbb{P}\ (DeltaReq \times dependency \times DeltaReq)$

$$\forall x, y, z: DeltaReq; a: dependency \mid (x, a, y) \in SIGSet$$
$$\bullet\ x\,.\,2 \in \{DelMT, UpdateMT\}$$
$$\Rightarrow x\,.\,1 \in TotalElem \wedge x\,.\,2 = AddMT$$
$$\Rightarrow x\,.\,1 \notin TotalElem$$
$$\wedge\ (x, a, x) \notin SIGSet$$
$$\wedge\ (y, a, x) \notin SIGSet$$
$$\wedge\ (y, a, z) \in SIGSet$$
$$\Rightarrow (x, a, z) \in SIGSet$$
$$AllSig = \{\ x: DeltaReq \mid \exists y: SIGSet \bullet\ x = y\,.\,1 \vee x = y\,.\,3\ \}$$

VI. Regression Testing Based on the Traceable Delta Model

In this section, the definitions of relevant concepts for platform independent testing are introduced.

Definition 5 (Unit of Testing Pattern). For a given system behavior in the form of a schema *TestModel*, a unit of testing pattern is a power set of *TestMMElem* elements.

$$UnitofTest: \mathbb{P}\ TestMMElem$$

Definition 6 (Abstract Test Case Pattern). For a given *TestModel*, an Abstract Test Case (ATC) pattern is a sequence of unit of testing patterns that made a path pattern. An ATC denotes a test requirement that should be traversed according to a selected coverage criterion.

$$ATC == \mathrm{seq}\ UnitofTest$$

According to an ATC definition, a Concrete Test Case (CTC) is a path pattern which proper values are assigned to its variables. The valid input space for a concrete test case is defined based on the input spaces of its variables; in our context, a valid input space for an ATC is a subset of the input variables which satisfy the corresponding guard conditions. It can be derived directly from the formal specification of an ATC.

Definition 7 (Debugging Traceability Link). A debugging traceability link is created between a behavioral model and test requirements at the platform independent level, in such a way that each behavioral model element can be a target for a test requirement in a coverage criterion. Debugging links encourage the process of locating faults using backward trace links. This type of link aims at solving debugging problems in a more precise way, e.g., finding elements that could have caused an observed fault and finding model elements that are dependent upon an erroneous element. Formally, a traceability link *TraceLink* is defined as a mapping between a predicate type and model elements to the requirement predicate *ReqPred* and the test meta-model *TestMMElem* as follows:

$$TraceLink: TestMMElem \times CovMetric \rightarrowtail \mathbb{P}\ ReqPred$$

DelatTraceLink: DeltaReq ↔ *ReqPred*

Definition 8 (Change Propagating Rule). To preserve consistency in agile model driven regression testing, it is required to propagate the delta model from the PIM to the PIT. Each transformation rule takes a delta model and two consistent models including a behavioral model and a test requirement model in a uniform format (e.g., XMI) and manipulates the latter such that consistency is held between them. For example, when an element is removed from a behavioral model by a delta operation DelMT, a rule should remove all ATCs that traverse the element. So after the rule execution, ATCs which traverse the deleted element don't exist in the test case pool (to keep safe access).

D. *Consistent regression test suites and dynamic selection*

The agile regression processes require the following, which is relevant for verification: firstly, being flexible, secondly, rapidly delivering verified working software as result of evolution. To provide such driven regression testing that covers agile changes of a model, known as safe technique, we classify initial ATCs into two categories: delta-traversing and non-delta-traversing ATCs.

The latter, called also *applicable* ATCs, don't traverse any delta model elements in their trace paths. This category is reusable for the new version of a model, but it is unnecessary to be rerun on the updated model; however, they are valid and reusable for regression testing of the future versions. The former visit at least one of the delta model elements in their trace paths and are divided into two distinct subsets of ATCs: *outdated* and *Retestable* ATCs.

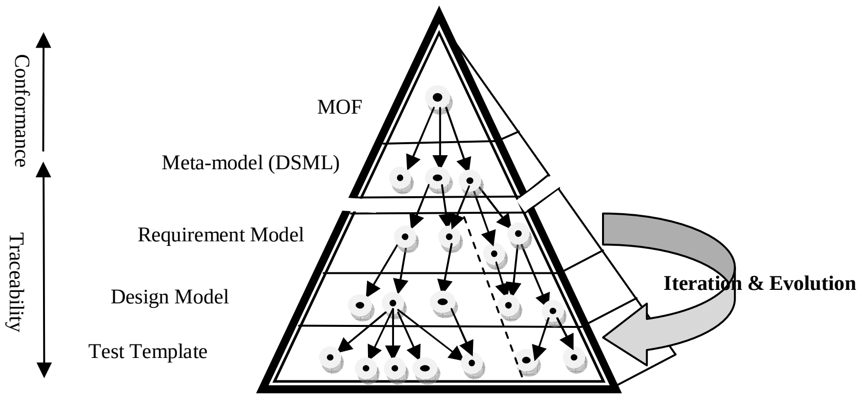


Figure 1. The traceability links in agile model driven regression testing

To categorize regression test cases, we need two parameters: *TotTrace* and *AllTest*. The former denotes all model elements that are traced by each ATC, i.e., the states, labels and transitions in a State Machine. The latter defines the original test suite for a behavioral model. It is clear that new ATCs are not a subset of *AllTest*.

AllTest: ℙ *ATC*
TotTrace: ATC → ℙ *TestMMElem*

According to these definitions, delta-traversing and non-delta-traversing test cases can be defined by the following schemas. The predicate parts of the schemas enforce that the elements of a delta-traversing ATC appear in delta signatures whose their *DeltaOp* is *DelMT* or *UpdateMT* while the intersection of the elements of non-delta-traversing ATCs and delta elements is empty.

Definition 9 (Retestable ATC). An *Retestable* ATC is a delta-traversing test case that traverses at least one delta signature which its delta operation is ‘Update’. *Retestable* ATCs should be re-executed in order to verify the correctness of an updated behavioral model. Formally:

RetestableATC

$\Xi Delta$
$Retestable!: \mathbb{P}\ ATC$

$Retestable! \subseteq AllTest!$
$\exists deltats: \mathbb{P}\ TestMMElem$
$\quad \mid deltats = \{\ e: TestMMElem \mid \exists s: AllSig \bullet\ s\,.\,1 = e \wedge s\,.\,2 = UpdateMT\ \}$
$\quad \bullet\ Retestable! = \{\ s: ATC \mid\ s \in AllTest! \wedge totalTrace!\ s \cap deltats \neq \emptyset\ \}$

Definition 10 (Outdated ATCs). An *outdated* ATC is a delta-traversing test case that traverses at least one delta signature which its delta operation is ‘Delete’. These ATCs are no longer valid to be rerun on the updated behavioral model. Formally:

OutdatedATC

$\Xi Delta$
$outdated!: \mathbb{P}\ ATC$

$\exists deltats: \mathbb{P}\ TestMMElem$
$\quad \mid deltats = \{\ e: TestMMElem \mid \exists s: AllSig \bullet\ s\,.\,1 = e \wedge s\,.\,2 = DelMT\ \}$
$\quad \bullet\ outdated! = \{\ s: ATC \mid\ s \in AllTest! \wedge totalTrace!\ s \cap deltats \neq \emptyset\ \}$

The remaining class consists of test cases that should be generated for testing the new functionalities of a model. This category is defined by the following schema:

NewATC

$\Xi Delta$
$newtest!: \mathbb{P}\ ATC$

$\exists deltats: \mathbb{P}\ TestMMElem$
$\quad \mid deltats = \{\ e: TestMMElem \mid \exists s: AllSig \bullet\ s\,.\,1 = e \wedge s\,.\,2 = AddMT\ \}$
$\quad \bullet\ newtest! = \{\ s: ATC \mid\ s \notin AllTest! \wedge totalTrace!\ s \cap deltats \neq \emptyset\ \}$

Finally, a safe regression test suite consists of Retestable and new test cases:

$RegressionTestCases\ \widehat{=}\ RetestableATC \vee NewATC$

According to the definitions, a detailed view of the framework is shown in Figure 2 including changes propagating rules and distinct categories of regression test suites after system refactoring. To verify the correctness of a modified design model, the corresponding delta model should be propagated from system specification to testing framework and a consistent test suite should be selected. In Figure 1, the traceability links in agile model driven regression testing is shown. Integrating the PIM and PIT in the approach, offers efficient traceability to derive suitable regression test suite and to keep consistency between software design and testing phase. It provides a typical infrastructure solution for expressing regression testing in the model driven fashion. *TestModel* and *TestModel’* denote that other behavioral models can be used in the approach if they are converted to a suitable model for software testing.

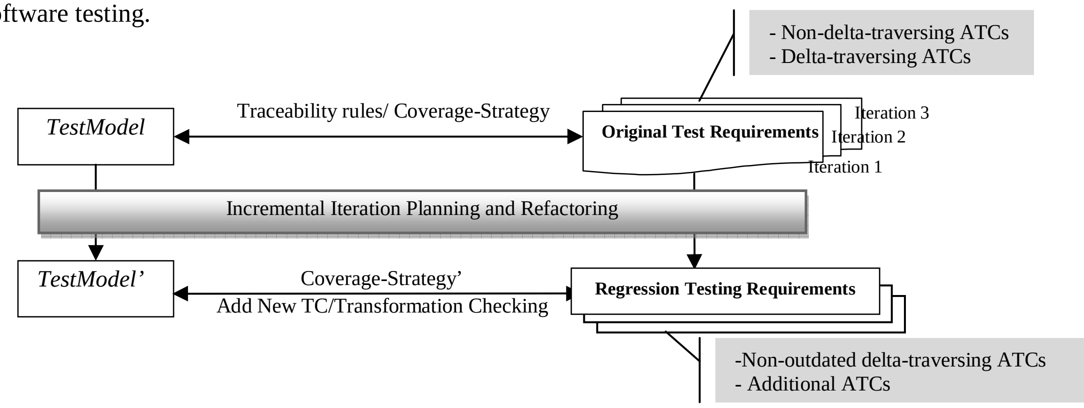


Figure 2. A detailed view of the framework.

VII. ON-THE-FLY AGILE TESTING

We introduce on-the-fly agile testing (also called online agile testing) for the maintenance phase which uses an incremental approach to generate and execute test requirements strictly simultaneously by traversing the new release of a model and validating changes. It can reduce time and memory since we no longer require an intermediate representation, such as the costly retesting of the specification using total test suites. Dynamic information from test execution can be incorporated in other heuristics, so that the regression testing process can adapt to the behavior of the system under evolution at runtime and therefore generate more revealing tests. Also, the coverage criteria can be strengthened for changed components to find more faults in changes parts of the model under test. When generating a test sequence, we can try to ignore as many test goals as possible in each state. This can provide an adaptive strategy for dynamic regression testing.

On-the-fly agile testing can also process changes by ranking them to track the actual choices they made in each iteration. The executed tests can also be recorded, yielding a regression test suite that can later be re-executed by common testing tools. The solution can consider prioritization metrics of coverage metrics as guidance while traversing the changed model, so that newly generated tests really raise the coverage criterion. In delta-based regression testing, one of the most important priority functions is the frequency of occurrence of delta elements in an ATC. It means that among regression test suites, ATCs that meet more delta elements in their traces need to be executed with a higher priority. It is advisable that each prioritization technique should be better than random prioritization and no-prioritization techniques. So, using prioritization approaches provides a support for agile change patterns while dealing with the complexity is one of the strengths of model-driven development tools. The main drawback of on-the-fly agile testing is its weak guidance used for test selection

Dealing with complex models, however, is almost exclusively an issue of the back end, using of filter functions and query-based approach to select the desired parts of an application model (code) will decrease the complexity of differencing techniques. In our work (at the automation phase) a novel class of transformations is used which are incrementally triggered by complex model change patterns. In the selected modeling environments, elementary model changes are reported on-the-fly by some live notification mechanisms to support undo/redo operations on the desired parts of systems. We describe a practical on-the-fly testing approach that uses ad hoc coverage criteria to produce and evaluate abstract test requirements. Each ad hoc coverage metric implies a delta-based query to cover specific changed/unchanged elements or delta dependencies in a behavioral model, e.g., coverage of states which updated in a new model or coverage of more than 3 changed elements. It can improve the quality of regression testing by providing meaningful test requirements in response to flexible queries. Complex queries can provide a narrow view aimed at covering specific parts in an updated model and extracting effective ATCs.

A delta-based coverage criterion *CovMetricType* is defined as a mapping between a behavioral model and a set of predicates *ReqPred* on the behavioral model. *CovMetricType* introduces two type coverage metrics: *StrategyBasedCC*, based on standard structural coverage criteria, and *AdHoc*, a new type of coverage metrics for agile regression testing .Coverage criteria satisfaction is defined by a total function from *CovMetricType* to a subset of abstract test cases. It ensures that there is at least one abstract test case to satisfy each predicate of *ReqPred*. Formally, it is defined as follows:

$$CovMetricType ::= StrategyBasedCC \mid AdHoc$$

$$Satisfication: CovMetricType \twoheadrightarrow \mathbb{P}\, ATC$$

$$\forall pred: ReqPred \bullet \{pred\} \in \operatorname{ran} CovMetric \Leftrightarrow (\exists_1 t: ATC \bullet \{t\} \in \operatorname{ran} Satisfication)$$

Definition 11 (on-the-Fly Agile Testing). An on-the-Fly Query Pattern (FQP) is built on the meta-model *TestMMElem* and the change tag *ChgTag* to investigate coverage of different elements that are changed by delta operations. An FQP is described by the recursive free type *FQP*. The *FQP* language enables testers to define significant queries to acquire adequate coverage of distinct test suites. Using *ChgTag* in the definition of FQP gives us a great opportunity to leverage synergies between testing and model transformation tools that work based on the labeled graph morphism.

$$EXP ::= and \langle\langle FQP \times FQP \rangle\rangle \mid or \langle\langle FQP \times FQP \rangle\rangle \mid not \langle\langle FQP \rangle\rangle \mid FQP^{+}$$

$$FQP ::= Traverse \langle\langle \mathbb{N} \times TestMMElem \times ChgTag \rangle\rangle \mid atomQuery \langle\langle EXP \times FQP \rangle\rangle \mid compQuery \langle\langle FQP \times EXP \times FQP \rangle\rangle \mid$$
$$QueryDep \langle\langle FQP \times dependency \times FQP \rangle\rangle$$

The end point of recursion is defined by the function *Traverse* to pass a tagged model element $\mathbb{N} \geq 1$ times. *atomQuery* defines negative queries and transitive closures FQP^{+}. A compound query *compQuery* enables developer to specify more complex queries by combining atomic queries using logical operation *AND* and *UNION*. An FQP can be extended to derive more meaningful test requirements in critical system testing which is beyond the scope of this paper. In complex queries a submodel or a combined pattern in a delta model can be investigated. To compare the coverage of different ad hoc coverage rules, we introduce the metric *coverage level* in Section 6. Also, The Filterfunc parameter of a FILTER expression enables to express constraints on the searched nodes or edges in the model which is queried.

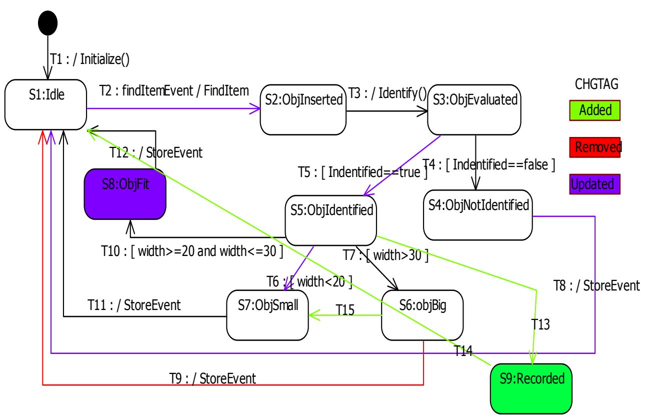
$$Term ::= Greater \langle\langle FQP \times FQP \rangle\rangle \mid Less \langle\langle FQP \times FQP \rangle\rangle \mid GreEq \langle\langle FQP \times FQP \rangle\rangle \mid LeEq \langle\langle FQP \times FQP \rangle\rangle$$
$$FilterFunc ::= Constant \mid Variable \langle\langle TestMMElem \rangle\rangle \mid Operation \langle\langle Term \rangle\rangle$$

Query driven regression testing provided by the extendable language syntax FQP in this definition will reduce the inherent complexity and the cost of evaluating complex design changes. Dealing with complex models, however, is almost exclusively an issue of the back end, using of filter functions to select desired parts, e.g., according to the fault diversity of the updated model or coverage-measures will decrease the complexity of regression testing. The flexible specification of the coverage criteria (standard and ad hoc) provides an interactive handle for defining accurate regression testing requirements. In order to obtain adequate coverage in regression testing, testers can use queries to specify effective coverage patterns at the model-level and to generate significant test suites incrementally.

VIII. Evaluation

E. *Case study on evolutionary constraints*

The most common way to apply logic coverage criteria to a state-based diagram is to consider the trigger of a transition as a predicate, then derive the logical expressions from the trigger. The example in Figure 3, inspired by [16], shows a finite state machine that models the behavior of a sorting machine at the design level. The optimized history of changes is shown in Listing. 1. The sorting machine accepts incoming objects depending on their size and fits them into suitable places. After undergoing corrections or improvements, changed elements are distinguished by the distinct tags. The main challenge is to determine a way to propagate the design changes using a well-defined delta model to the original test suites and to provide a technique based on the delta model to select an efficient and consistent regression test suite. The event message used is this example is signal event and the guard condition is true for all transitions, which means the transitions will be triggered when the specified signals are received on the port. Thus, the logical expressions derived from the triggers all consist of at least a clauses. Take transition 10 for example, if the guard is false, the predicate should be ¬ (width >= 20 and width <= 30).



S8: (*State, UpdateMT*,0, OldValue, NewVal)
S9: (*State, AddMT*, 0,empty, NewVal)
T2: (*Lable, UpdateMT*, OldValue, NewVal)
T5: (*Lable, UpdateMT*, OldValue, NewVal)
T6: (*Lable, UpdateMT*, OldValue, NewVal)
T8: (*Lable, UpdateMT*, OldValue, NewVal)
T13: (*Transition, AddMT*, empty, NewVal)
T14: (*Transition, AddMT*, empty, NewVal)
T15: (*Transition, AddMT*, empty, NewVal)
T9: (*Transition, DelMT*, OldVal, empty)

Figure 3. The finite state machine of a sorting machine

Listing 1. The optimized delta model of Figure 3

In order to evaluate the effectiveness of the delta-based coverage criteria, we introduce coverage level of a delta-based criterion. Each updated or new element in a delta operation may be covered by a regression test requirement or not, which denoted by the binary decision variable bi ∈ {0,1}. The parameter di denotes a non-removed delta element i in a specific coverage criterion (e.g., in delta-predicate coverage rule, di denotes a changed predicate i) and, |Tot| determines the number of non-removed changed elements in a behavioral model. We define Coverage level (Cov) of a delta-based criterion C as follows:

$$Cov\,C = \% \frac{\sum_{i=0}^{n} b_i d_i}{|Tot|} \quad (1)$$

If the value of *Cov* for a delta-based coverage criterion equals to '1' then the selected regression test suite is 100% safe. The purpose of calculating *Cov* is measuring of structural coverage and the effect of program and model structures on logic-based test adequacy coverage. It is a serious need for coverage criteria measuring irrespective of implementation structure, or a technical way of structuring test plan with focus on adequate coverage. Using this metric, testers can compare different coverage criteria. Obviously, equivalent coverage criteria in terms of the coverage level should have the same value. Since ad hoc coverage criteria (as customized reduction techniques) select a representative subset from the original test pool, it is vital to evaluate the adequacy of the resulting subset using the coverage level metric and to compare its value by proper assigned values.

We calculate the coverage level for some coverage rules on the case study in Table 2. The edge set after the changes is denoted by vector <T1, T2, T3, T4, T5, T6, T7, T8, T9, T10, T11, T12, T13, T14, T15>. As shown in Table 2, the analysis of reduction percentage for two rules with the same *Cov* can provide a strong indicator for testers to keep the size of test suites according to the restricted testing resources.

Table II
THE COVERAGE LEVEL ANALYSIS FOR SOME COVERAGE RULES

Coverage Rules	**Binary Coverage Vector**	***Cov***
Delta-transition coverage	<0,1,0,0,1,1,0,1,-,0,0,0,0,1,1,1>	100%
Additional-transition coverage	<0,0,0,0,0,0,0,0 ,-,0,0,0,0,1,1,1>	42.86%
Updated-transition coverage	<0,1,0,0,1,1,0,1,-,0,0,0,0,0,0,0>	57.14%
Additional-clause coverage	<0,0,0,0,0,0,0,0 ,-,0,0,0,0,1,1,1>	42.86%
Updated-predicate coverage	<0,1,0,0,1,1,0,1,-,0,0,0,0,0,0,0>	57.14%
Coverage of at least two updated predicates	<0,0,0,0,0,0,0,1,-,0,0,0,0,0,0,0>	14.28%
	<0,1,0,0,0,1,0,0,-,0,0,0,0,0,0,0>	28.57%
Coverage of at least four additional predicates	<0,0,0,0,0,0,0,0 ,-,0,0,0,0,0,1,1>	28.57%
	<0,0,0,0,0,0,0,0 ,-,0,0,0,0,1,0,1>	28.57%
	<0,0,0,0,0,0,0,0 ,-,0,0,0,0,1,1,1>	42.86%

F. Reduction analysis

We apply our methodology on two case studies including the behavior of a Personal Investment Management System (PIMS) [17] and data transfer interface of computing device [18] which underwent a major design change. PIMS aims a person who has investments in banks and stock market for book keeping and computations concerning the investments using software assistance. A data transfer interface of computing device describes the behavior of device in data transfer rate, power consuming and management, full duplex data communications, interrupt driven functionality management.

There are three main differences between these two case studies: (1) Case study B is smaller both in terms of the model size (number of elements) and the test suite size (number of ATCs generated for the input model); (2) The number of faults detected by the test suite is much higher for case study B; (3) The fault detection in case study A depends on the input data, whereas ATCs in case study B either detect or not a fault regardless of input data; and (4) The change rate is much higher for case study A.

In our experimental studies, we investigate the number of test cases after changes in the retest-all selection, DbRTS, optimized DbRTS, random-based selection and Similarity-based Test Case Selection (STCT) [19] and [20] techniques after modifying the system model. The optimized DbRTS method removed redundant modification-traversing test case from consistent test suites provided by the DbRTST. It can use the prioritization techniques (e.g., frequency of occurrence of changes or a longer path) to remove redundant ATC among different test cases which traverse a specific delta element. An STCS is composed of: similarity matrix generation where each matrix cell represents the similarity value for a pair of well-defined ATCs based on the given similarity function, and reducing ATCs where an optimization algorithm selects a subset of the original ATCs with minimum sum of similarities.

Figure 4 shows the boxplots of the variant size of regression suite when considering on three traditional releases of the case studies over 100 runs. Figure 4.a and Figure 4.b show the same boxplots for case studies A and B, respectively. After performing some TCS algorithms as shown in Figure 4, each test selection size of a variant has a score in the range (0, 250). As visible from the boxplots and confirmed by the statistical tests, each algorithm has a significant effect on ranks.

The most clear result that the figure conveys is that there is a common pattern between the three releases. As shown in the comparison result of Figure 4.a and 4.b, using the optimized DbRTST leads to a reasonable reduction in re-executing of all test cases. In fact, optimized DbRTST always performs better than retest-all, and better than or equal to DbRTS and STCT while applied in the same range. It proposes the relative effectiveness of the evaluated TCS techniques for regression testing in an agile software development context, especially, when the development moves toward rapid-releases

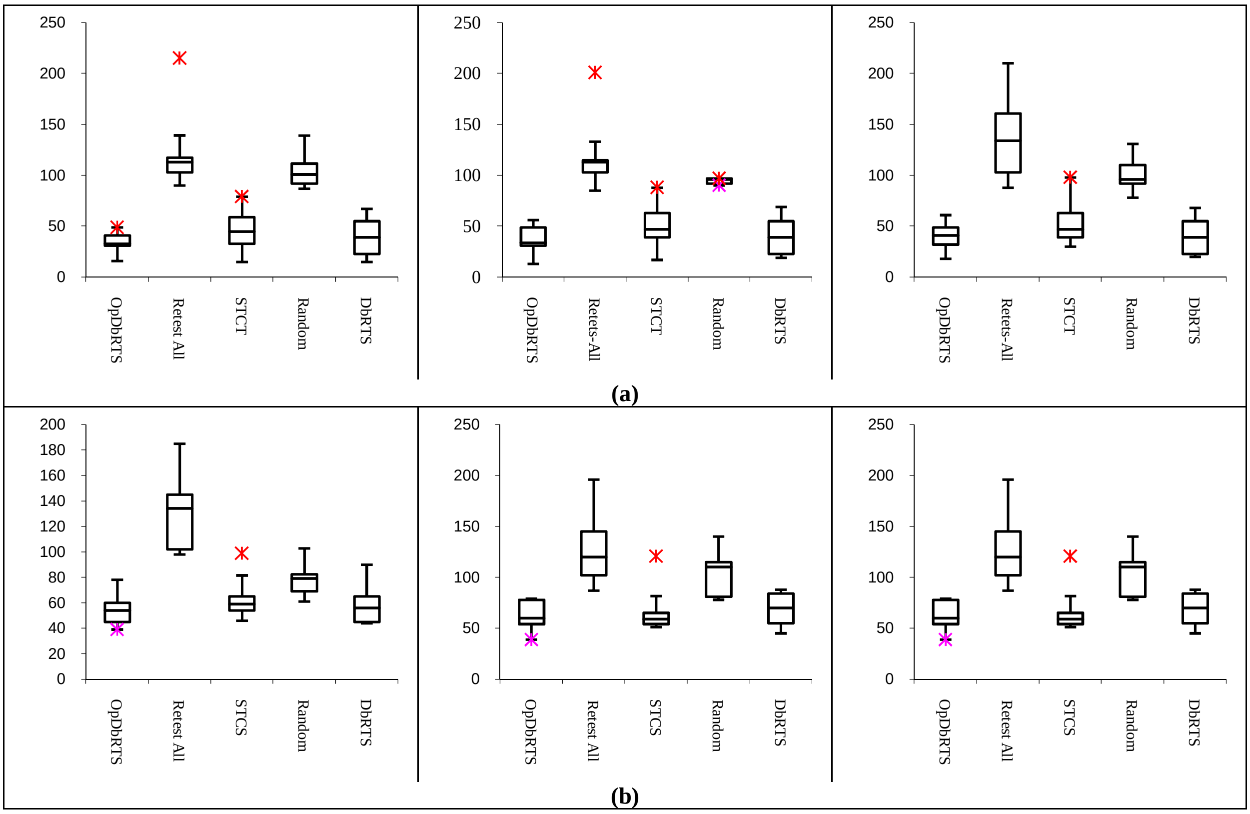


Figure 4. The ranking of test suite size for five selection algorithms.

IX. Related work

Traditionally, there are two methods for specification-based regression testing, including formal and informal methods. Although integrated strategies by combining the advantages of UML and formal methods have attracted increasing attention, there is still a lack of integrated approaches in regression testing. The main benefits of applying abstract testing to agile development have been explained in [21]; e.g., meaningful coverage, more flexibility, low maintenance costs. The studies in related literature don't provide results for reduction rate, change and fault coverage and on-the-fly testing, especially on industrial case studies. The following subsections present samples of research conducted in each of these areas.

G. *Model-based on informal methods*

UML as a modeling language is designed to provide a standard way to visualize the design of a system. Various UML-based regression testing techniques are reported in the different substantial research areas of regression testing, e.g., [7], [22], [23], [24], [2] and [25] in UML-based RTSs; and [19], [20] and [26] in UML-based regression test prioritization and minimization techniques.

A survey of RTSs provided by Rothermel and Harrold [22] and a recent systematic review is presented by Engström et al. [27] that classified model-based and code-based RTSs. Briand et al. Briand et al. [2] proposed a RTST based on analysis of UML sequence and class diagrams. Their approach adopts traceability between the design model(s), the code and the test cases. They also present a prototype tool to support the proposed impact analysis strategy.

Farooq et al. [28] presented a RTS approach based on identified changes in both the state and class diagrams of UML that used for model-based regression testing. They utilized Briand et al. [2] classification to divide test suites into obsolete, reusable and re-testable. Also, an Eclipse-based tool for model-based regression testing compliant with UML2 is proposed in their research.

Chen et al. [24] proposed a specification-based RTS technique based on UML activity diagrams for modeling the potentially affected requirements and system behavior. They also classified the regression test cases that are to be selected into the target and safety test cases based on the change analysis. Wu and Offutt [25] presented a UML-based technique to resolve problems introduced by the implementation transparent characteristics of component-based software systems. In corrective maintenance activities, the technique started with UML diagrams that represent changes to a component, and used them to support regression testing. Also, a framework to appraise the similarities of the old and new components, and corresponding retesting strategies provided in this paper.

Tahat et al. [26] presented and evaluated two model-based selective methods and a dependence-based method of test prioritization utilizing the state model of the system under test. These methods considered the modifications both in the code and model of a system. The existing test suite is executed on the system model and its execution information is used to prioritize tests.

Some UML-based approaches cover MDA aspects are: Naslavsky et al. [29] presented an idea for regression testing using class diagrams and sequence diagrams using MDA concepts. They make use of traceability for regression testing in the context of UML sequence and class diagrams. Farooq [30] discussed a model driven methodology for test generation and regression test selection using BPMN 2.0 and UML2 Testing Profile (U2TP) for test specification while a trace model is used to express relation between source and target elements.

Some studies work on integrating of MBT and AD e.g.: [31] uses AD to improve MBT and also MBT within AD. It proposes MBT outside the AD team, i.e., not strongly integrated. Ref. [32] aimed to adapt MBT for AD and also suggests using MBT within the AD principles, but does not propose in detail how to modify AD for productive integration, e.g., adapting specifications. Ref. [33] used a restricted domain and a limited property language (weaker than the usual temporal logics). It uses very strict models that are lists of exemplary paths. A MBT approach prescribed by Faragó [21] as a specific approach to MBT with limited applicability, whereas we provide a more general viewpoint and discuss how MBT may be applied to agile/lightweight methodologies using testing techniques based on model transformations.

Pilskalns et al. [34] discussed another modeling approach for regression testing the design models instead of testing the concrete code. In some studies, similar DSMLs are used for regression testing, e.g., Yuan et al. [35] utilized the business process diagram and transformed it into an abstract behavioral model to cover structural aspects of test specification, and [23] presented an approach for model-based regression testing of business processes to analyze change types and dependency relations between different models such as Business Process Modeling Notation (BPMN), Unified Modeling Language (UML), and UML Testing Profile (UTP) models. A way of handling structural coverage analysis compared is described in Kirner [36]. The author provides an approach to ensure that the structural code coverage achieved at a higher program representation level is preserved during transformation down to lower program representations. The approach behind testability transformation described by Harman et al. [37] is a source to source transformation. The transformed program is used by a test data generator to improve its ability to generate test data for the initial program. Baresel et al. [38] provided an empirical study, the relationship between achieved model coverage and the resulting code coverage. They describe experiences from using model coverage metrics in the automotive industry. The conclusion of the experiment is that there are comparable model and code coverage measurements, but they heavily depend on how the design model was transformed into code. The research reported in [10] proposed an approach for using UML models for usual testing purposes. Models are used as test data, test drivers, and test oracles. UML Object diagrams are proposed to be used as test data and oracles by exhibiting the initial and also final conditions. The values of concrete attributes related to testing are specified in these diagrams.

The spread of UML-based testing methods has led to the creation of the UML Testing Profile (UTP) as an OMG standard [39]. UTP proposes extensions to sequence diagrams, such as test verdicts and timing issues, which allow these diagrams to be used as bases for test case generation. As noted in [40], code generation from behavioral models automatically, as is usual in model-driven engineering approaches, allows the concept of TDD to be applied at a higher level of abstraction. It defines only structural models and leaves method bodies to be developed by programmers.

The agile/lightweight technique discussed in [41] allows proposes an extension to object diagrams to define positive/negative object configurations that the corresponding class diagram should allow/disallow. It is possible to consider each “modal object diagram” as a test case that the corresponding class diagram should satisfy. A fully automated technique based on model-checking performs the verification on abstract models.

In the research reported in [42] a program is discussed that processes the models themselves. The testing of such applications is investigated, and a technique based on the use of meta-models is proposed. It is mentioned that precise defined meta-models allow the automatic or semi-automatic creation of simple test cases. The approach presented in [43] proposes a method which allows further automation of this technique, by composing test cases from valid, manually created "model fragments". Ref. [44] attempts to define viable mutation operators which function based on the knowledge contained in the meta-model. It also shows how these operators can be customized for specific domains. It generalizes the technique to be applied both to domain-specific meta-models and technical models created during the software lifecycle. Ref. [45] describes a model based testing theory in an agile manner where models are expressed as labeled transition systems. Definitions, hypotheses, an algorithm, properties, and several examples of input output conformance testing theories are discussed.

H. Model-based on formal methods

Formal methods, based upon basic mathematics, provide an unambiguous complement for validating and verifying software artifacts at an appropriate level of abstraction. Software testing based on formal specifications can improve the software quality through early detection of specification errors. Different approaches are carried out for test case generation using formal methods, especially Z, for example [46], [47], [48] and [49]. But there are scarcely relevant works in the field of regression testing that support formal specifications.

To the best knowledge of the authors, there is not an integrated method of MDA and formal notations for agile regression testing. The proposed formal framework, not only can extend to support code generation in MDD using refinement technique, but also provides the executable semantics to the modeling notations which is a shortage in UML2 + OCL notations.

Our approach for model driven regression test selection is safe, efficient and potentially more precise than the other similar approaches. DbRT by means of exploring fine-grained traceability links, leveraging on-the-fly testing and supporting the direct analysis of on-the-fly patterns to determine affected elements is capable of achieving better precision than the similar model-based approaches. The approaches provided in [2], [28], [25] and [35] support model-based regression test selection count on traceability relationships among artifacts. These selective approaches are safe, efficient and precise. But, the majority of the approaches do not support on-the-fly testing, automatic refinement and fault modeling in model-based regression testing. Query driven regression testing and the ability of covering customized evolving parts in a model, as the new capabilities in regression testing, are proposed in this paper. Finally, our approach due to its formalizing style can be extended to other DSMLs while the domain of the other approaches is limited to a specific modeling language.

X. CONCLUSION AND FUTURE WORK

Change management has always been a challenge in software development, whether you use agile *methods* or not. If changes are needed, in agile, they can be recognized earlier and interleaved with earlier iterations. This also provides a smooth way to also get development risks out of the way, earlier in the development cycle. Rapid change management is the rationale for using agile methods and can add significant value to resulting software. MDA is the next logical evolutionary step to complement 3GLs in the business of software engineering. The main shortcomings of them (according to our model driven approach) is supporting of the essential concepts of MDD e.g., change management, traceability and refinement. As an example, we need to transform a platform independent model to a platform specific model which is considered as a weakness of these tools.

The approach aims to keep pace with agile model driven regression testing in a formal way. The framework can solve a number of restricted outlooks in model driven regression testing, including: well-defined delta signature definition; regression test suite identification and some limiting factors of incremental maintenance and retesting. In this paper, key concepts of agile model driven regression testing includes system refactoring, model transformation, traceability, delta-based coverage criteria are formalized by Z-notations that can provide the input of Z-based tools for various kinds of dynamic and static analysis and functional verification.

We divide abstract test suites into distinct categories according to the delta operations which they traverse. Test case generation for new functionalities of a system is performed by applying derivation rules on recently added elements and targeting them in new coverage rules. In addition, a precise formalism to define model-level coverage criteria and adequate test requirements is proposed.

An issue for future work is important differences for identifying the logical model coverage for agile regression testing when: (i) other DSMLs may be use as modeling languages, and therefore implementation styles, (ii) The modeling language may use a high level of abstraction which complicates the identification of logical model coverage metrics. (iii) Many modeling environments are considered as development platforms.

REFERENCES

[1] D Thomas, "MDA: Revenge of the modelers or UML utopia?," IEEE Software, pp. 21 (3): 15–17, 2004.

[2] L C Briand, Y Labiche, and S He, "Automating regression test selection based on UML designs," Information & Software Technology, vol. 51, no. 1, pp. 16-30, 2009.

[3] J M Spivey, The Z Notation: A Reference Manual, Second edition ed.: Prentice Hall, Englewood Cliffs, NJ, 1992.

[4] P Malik and M Utting, "CZT: A Framework for Z Tools, Formal Specification and Development in Z and B," in Lecture Notes in Computer Science 3455, 2005, pp. 62-84.

[5] M Saaltink, "The Z/EVES System," in In Proceedings of the 10th International Conference of Z Users on The Z formal Specification Notation, Lecture Notes in Computer Science 1212, 1997, pp. 72-85.

[6] S Yoo and M Harman, "Regression testing minimization, selection and prioritization: a survey," Softw. Test., Verif. Reliab, vol. 22, no. 2, pp. 67-120, 2012.

[7] G Rothermel, R H Untch, C Chu, and M J Harrold, "Prioritizing Test Cases For Regression Testing," IEEE transactions on software engineering , vol. 27, no. 10929-948, 2001.

[8] R. C. Martin, Agile software development: principles, patterns, and practices.: Prentice Hall PTR, 2003.

[9] K. Beck, Test Driven Development: By Example.: Addison-Wesley Professional, 2002.

[10] B Rumpe, "Model-based testing of object-oriented systems ," in in Formal Methods for Components and Objects, International Symposium, FMCO, ser. Lecture Notes in Computer Science, vol. 2852. , 2002, pp. 380–402.

[11] J H Hausmann, R Heckel, and r S Saue, "Extended model relations with graphical consistency condition," in UML Workshop on Consistency Problems in UML-based Software Development , 2002, pp. 61-74.

[12] P Stevens, "Bidirectional model transformations in QVT: semantic issues and open questions," Softw. Syst. Model , vol. 9, no. 1, pp. 7-20, 2010.

[13] R Heckel, J M Kuster, and G Taentzer, "Confluence of typed attributed graph transformation systems. In: Graph Transformation," in First International Conference. Lecture Notes in Computer Science 2505, 2002, pp. 161–176.

[14] T Mens, "A State-of-the-Art Survey on Software Merging. ," IEEE Trans. Softw.Eng, vol. 28, no. 5, pp. 449–462., 2002.

[15] G Bergmann, I Ráth, G Varró, and D Varró, "Change-driven model transformations," Software & Systems Modeling, vol. 11, no. 3, pp. 431-461, 2012.

[16] S Weißleder, Test Models and Coverage Criteria for Automatic Model-Based Test Generation with UML State Machines. Humboldt-Universität zu Berlin.: PhD Thesis, 2009.

[17] P Jalote, An Integrated Approach to Software Engineering, 3rd ed.: Springer , 2006.

[18] R B'Far and R Fielding, Mobile Computing Principles: Designing and Developing Mobile Applications with UML and XML.: ISBN-13: 978-0521817332, 2004.

[19] H Hemmati, A Arcuri, and L Briand, "Achieving scalable model-based testing through test case diversity," ACM Trans. Softw. Eng. Methodol, vol. 22, no. 1, p. 6 pages, 2013.

[20] H Hemmati, A Arcuri, and L Briand, "Reducing the cost of model-based testing through test case diversity," in Proceedings of the 22nd International Conference On Testing Software and Systems. Lecture Notes in Computer Science 6435, 2010, pp. 63–78.

[21] D Faragó, "Model-based testing in agile software development," in in 30. Treffen der GI-Fachgruppe Test, Analyse & Verifikation von Software (TAV), Testing meets Agility, ser.Softwaretechnik-Trends, 2010.

[22] G Rothermel and M J Harrold, "Analysing Regression Test Selection Techniques," IEEE Transactions on Software Engineering, vol. 22, no. 8, pp. 529-551, 1996.

[23] Q Farooq, S Lehnert, and M Riebisch, "Analyzing Model Dependencies for Rule-based Regression Test Selection," in Modellierung, 2014, pp. 19-21.

[24] Y Chen, R Probert, and D Sims, "Specification-based regression test selection with risk analysis," in Proceedings of the 2002 conference of the Centre for Advanced Studies on Collaborative research, 2002, p. 1.

[25] Y Wu and J Offutt, "Maintaining Evolving Component based software with UML," in Proceeding of 7th European conference on software maintenance and reengineering, 2003, pp. 133-142.

[26] L Tahat, B Korel, Harman M, and H Ural, "Regression test suite prioritization using system models," Softw. Test., Verif. Reliab, vol. 22, no. 7, pp. 481-506, 2012.

[27] E Engström, M Skoglund, and P Runeson, "Empirical evaluations of regression test selection techniques: a systematic review. Proceedings of the Empirical software engineering and measurement," in ACM-IEEE international symposium , 2008, pp. 22–31.

[28] Q Farooq, "A Model Driven Approach to Test Evolving Business Process based Systems," in 13th International Conference on Model Driven Engineering Languages and Systems, 2010, pp. 6-24.

[29] L Naslavsky and D J Richardson, "Using Traceability to Support Model-Based Regression Testing," in Proceedings of the twenty-second International Conference on Automated Software Engineering, 2007, pp. 567-570.

[30] Q Farooq, Z M Iqbal, ZI Malik, and A Nadeem, "An approach for selective state machine based regression testing," in Proceeding of AMOST ACM , 2007, pp. 44–52.

[31] M Utting and B Legeard, Practical Model-Based Testing: A Tools Approach. San Francisco, CA, USA: Morgan Kaufmann , 2007.

[32] O P Puolitaival, "Adapting model-based testing to agile context," in ESPOO, 2008.

[33] M Katara and A Kervinen, "Making Model-Based Testing More Agile: A Use Case Driven Approach. ," in In Eyal Bin, Avi Ziv, and Shmuel Ur, editors, Haifa Verification Conference, 2006.

[34] O Pilskalns, G Uyan, and A Andrews, "Regression Testing UML Designs," in ICSM. 22nd IEEE International Conference on , 2006, pp. 254-264.

[35] Q Yuan, J Wu, C Liu, and L Zhang, "A model driven approach toward business process test case generation," in 10th International Symposium on Web Site Evolution, 2008, pp. 41-44.

[36] R Kirner and S Kandl, "Test coverage analysis and preservation for requirements-based testing of safety-critical systems," ERCIM News , vol. 75, pp. 40–41, 2008.

[37] M Harman et al., "Testability transformation," Software Engineering, IEEE Transactions , vol. 30, no. 1, pp. 3-16, 2004.

[38] A Baresel, M Conrad, S Sadeghipour, and J Wegener, "The interplay between model coverage and code coverage," in EuroSTAR, 2003.

[39] P Baker, Z R Dai, O Haugen, and I Schieferdecker, Model-Driven Testing: Using the UML Testing Profile.: SPRINGER, 2008.

[40] G Engels, B Gldali, and M Lohmann, "Towards model-driven unit testing," in in MoDELS Workshops ser. Lecture Notes in Computer Science, 2006, pp. 182–192.

[41] S Maoz, J O Ringert, and B Rumpe, "Modal object diagrams," in in European Conference on Object- Oriented Programming, ECOOP, ser. Lecture Notes in Computer Science,vol. 6813, 2011.

[42] J Steel and M Lawley, "“Model-based test driven development of the Tefkat model-transformation engine," in in International Symposium on Software Reliability Engineering, ISSRE. , 2004, pp. 151–160.

[43] E Brottier, B Fleurey, J Steel, B Baudry, and Y L Traon, "Metamodel-based test generation for model transformations: an algorithm and a too," in in International Symposium on Software Reliability Engineering, ISSRE, 2006, pp. 85–94.

[44] J Mottu, B Baudry, and Y L Traon, "Mutation analysis testing for model transformations ," in in European Conference on Model Driven Architecture, 2006, pp. 376–390.

[45] J Tretmans, "Model Based Testing with Labelled Transition Systems and Testing," Formal Methods, pp. 1–38, 2008.

[46] H Liang, "Regression Testing Of Classes Based On TCOZ Specifications," in. Proceedings of the 10th International Conference on Engineering of Complex Computer Systems , 2005, pp. 450-457.

[47] B Mahony and J S Dong, "Timed Communicating Object Z," IEEE Transactions on Software Engineering, vol. 26, no. 2, pp. 150-177, 2000.

[48] J A Jones and Harrold, M J , "Test-suite reduction and prioritization for modified condition / decision coverage," IEEE Transactions on Software Engineering, vol. 29, no. 3, pp. 195–209., 2003.

[49] M Cristià and P R Monetti, "Implementing and applying the Stocks-Carrington framework for model-based testing," in Springer-Verlag ICFEM 5885, 2009, pp. 167–185.

COMPARATIVE ANALYSIS OF EARLY DETECTION OF DDoS ATTACK AND PPS SCHEME AGAINST DDoS ATTACK IN WSN

Kanchan Kaushal, M.Tech Student

Varsha Sahni, Assistant professor

Department of Computer Science Engineering, CTIEMT Shahpur Jalandhar, India

ABSTRACT- **Wireless Sensor Networks carry out has great significance in many applications, such as battlefields surveillance, patient health monitoring, traffic control, home automation, environmental observation and building intrusion surveillance. Since WSNs communicate by using radio frequencies therefore the risk of interference is more than with wired networks. If the message to be passed is not in an encrypted form, or is encrypted by using a weak algorithm, the attacker can read it, and it is the compromise to the confidentiality. In this paper we describe the DoS and DDoS attacks in WSNs. Most of the schemes are available for the detection of DDoS attacks in WSNs. But these schemes prevent the attack after the attack has been completely launched which leads to data loss and consumes resources of sensor nodes which are very limited. In this paper a new scheme early detection of DDoS attack in WSN has been introduced for the detection of DDoS attack. It will detect the attack on early stages so that data loss can be prevented and more energy can be reserved after the prevention of attacks. Performance of this scheme has been seen by comparing the technique with the existing profile based protection scheme (PPS) against DDoS attack in WSN on the basis of throughput, packet delivery ratio, number of packets flooded and remaining energy of the network.**

KEYWORDS

DoS and DDoS attacks, Network security, WSN

1. INTRODUCTION

Network security is one of the major issues emerging now a days and catch people's attention. Distributed Denial of Service (DDoS) attacks is major threat to internet today.

Distributed denial-of-service attacks (DDoS) can be expressed by an immense number of packets being sent from numerous attack sites to a fatality site. These packets appear in such a high quantity that some of the key resources at the fatality (buffers, bandwidth, CPU time to evaluate responses) are exhausted quickly. The fatality either crashes or takes so much time to handle the attack traffic that it can-not give attention to its real work. So authorized clients are underprivileged of the service of victim for as long as the attack lasts.

Denial-of-service (DoS) and distributed-denial-of-service (DDoS) attacks are becoming dangerous to Internet operation. They are, basically, resource overloading attacks. The intent of the attacker is to link up a selected key resource at a victim, most often by sending high volume of apparently legitimate traffic that requests some of the services from the victim. The over utilization of resources cause humiliation or denial of the victim's service to its authorized clients. The major difference between DoS attack and

DDoS attacks is in its scale. DoS attack uses only one attack machine to generate malicious traffic whereas DDoS attacks use more than one attack machines. Both of these attacks consume the limited and useful resources in the network leads to increase in energy consumption, delay in response and decreased throughput. Many of the defensive mechanisms has already been proposed to mitigate the effect of DoS and DDoS attacks. Many of these defensive techniques works after the attack has already been launched into the network which leads to loss of data packets and more consumption of energy. So, we have established a new approach for the detection of DDoS attacks that will detect the attack on early stages before it is completely launched into the network and helps to reserve more energy, prevents loss of data and gives increases throughput. In this paper we have verified the results by comparing the new approach with the existing Profile based protection scheme (PPS) for the detection of DDoS attacks and the new scheme is giving better results than the PPS scheme in case of throughput, Energy reserved, Packet Delivery Ratio and Flood count.

The rest of the paper is organized as follows: in section 2 the related work has been discussed. Section 3 discusses the motivation and points out the derivation of the recent works. The section 4 gives the detail of the proposed work. In section 5, simulations are discussed and different parameters are analyzed with their plots. Then section 6 discusses the conclusion of proposed technique.

2. RELATED WORK

[1] provides a scheme that check the profile of each node in network and only the attacker is one of the node that flood the network with unnecessary packets then PPS has block the performance of attacker. The simulation results represent the same performance in case of normal routing and in case of PPS scheme; it means that the PPS scheme is effective and it shows 0% infection in existence of attacker.

In [2]; two types of attacks on WSN that are jamming and flooding has been discussed and this paper provides an efficient technique to detect jamming and flooding attacks. The method discussed in this paper provides improved performance over the existing methods.

Article [3] examines how attacks happen in WSNs and differentiate these attacks by conducting a survey. However, the main aim of this analysis is to examine how to prevent such attack in the WSNs by creating a sound understanding about various kinds of attacks in WSNs.

[4] has explored the WSN architecture according to the OSI model with some protocols in order to achieve good background on the WSNs and help readers to find a summary for ideas, protocols and problems towards an appropriate design model for WSNs.

In paper [5], the authors proposed a novel IDS based on energy prediction (IDSEP) in cluster-based WSNs. The main concept of IDSEP is to recognize hostile nodes on the basis of energy consumed by the sensor nodes. Sensor nodes with abnormal energy consumption are marked as malicious ones. Besides, IDSEP is depicted to differentiate classes of ongoing DoS attacks on the basis of energy consumption thresholds. The simulation results show that IDSEP detects and recognizes hostile nodes effectively.

In [6] Authors are conducting a review on DDoS attack to show its impact on networks and to present various defensive, detection and preventive measures adopted by researchers till now.

[10] Shows that the absence of central monitoring unit makes it vulnerable to various attacks. Denial of service attack (Dos) is an active internal attack which degrades the performance WSN. This attack can be distributed in nature on the basis of intent of attack. In this paper, authors uses modified variant of Ad-hoc On Demand Distance Vector (AODV) protocol to examine the consequences of Dos attack on performance of system and then apply the prevention technique to examine the change in performance of network.

[13] Traditional security schemes developed for sensor networks are not satisfactory for cluster-based WSNs because of their vulnerability to DoS attacks. In this paper, authors proposed a security scheme against DoS attacks (SSAD) in cluster-based WSNs. This technique organizes trust management with energy character that allows nodes to choose the trusted cluster heads. Besides, a new type of vice cluster head node is proposed to detect malicious cluster heads. The analyses and simulation results proves that SSAD can detect and prevent betrayed nodes successfully.

3. MOTIVATION

WSN consists of various numbers of nodes. Each node in WSN have limited amount of energy and resources. Energy conservation is the main focus in every research of WSNs. The nodes in wireless sensor network communicate wirelessly with each other, the wireless nature makes them susceptible to various kinds of attacks such as black hole attack, worm hole attack , denial of service attack, DDoS attack etc. While attacks such as black hole or wormhole focus on the loss of the data, the distributed denial of service attacks consisting of more than one attacker node, focuses on consumption of resources of the network such as bandwidth and energy of the nodes.

When the source node has to send data to the destination node in the network, it broadcasts route request messages in the network to its one hop neighbor nodes which are in its radio range. The nodes upon receiving the route request replies back to the source if they have an route to the destination node, else they re-broadcast the request message to their own neighbors. The process is continued until the request reaches the destination, however if there are malicious nodes present in the network with the intention of consuming the network resources, upon receiving the request they floods the network with such messages. This consumes maximum bandwidth, resources and energy of the network.

In the study done by the authors in [1] they have detected the malicious nodes in the network using the profile based scheme which relies on analyzing the behavior of the nodes in the network. The past pattern of the nodes is compared with the current patterns, any abnormality leads to detection of the malicious node in the network. In such cases the abnormality arises upon the occurrence of the attack leading to consumption of the resources.

Hence there must be a need to detect the malicious nodes in the network at the early stages so that resources of the network are minimized. In our proposed protocol, the attacker will be detected at the early stages before it is completely launched into the network that will prevent the loss of data in the network, will reserve more energy after the effect of attack, reduces the flood count in the network and enhance the throughput.

4. SIMULATION PARATMETERS

The simulation is implemented In Network Simulator 2.31, [11]. The simulation parameters are provided in Table 1. We implement the random waypoint movement model for the simulation, [11] in which a node starts at a random position, the simulation time is 30 seconds, and radio range is 250 meters. A packet size of 512 bytes and has cbr/udp type of traffic. Type of attack is DDoS attack and number of attacker will vary from 1 to 4.

TABLE 1: Simulation Parameters

Parameters	Values
Examined Protocol	AODV
Number of nodes	56
Simulation time (in seconds)	30
Dimensions of simulated area	1000*1000
Traffic type	Cbr/udp
Radio range	250mtrs
Types of attack	DDoS
Packet size (in bytes)	512
DDoS attacker nodes	1,2,3,4

5. THE PROPOSED TECHNIQUE

In this section, we present our proposed method Early detection of DDoS attacks in WSN in which the attacker is identified on the basis of the number of transmissions corresponding to the number of neighbors of a node and these transmissions are compared with the threshold value computed and PDR of other nodes in the network. Description of early detection of DDoS attack in WSN is given in the following subsections.

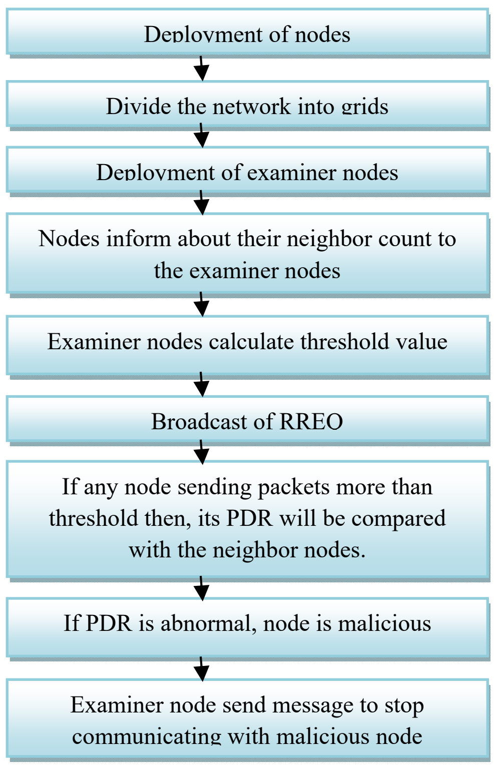


Fig 1: Flowchart for working of proposed protocol

5.1 Deploy nodes and Grid formation

First step is to deploy nodes in the specified region of 1000*1000 sq. mts. 50 nodes has been deployed in this region with radio range of 250 mtrs. Then the whole network will be divided into grids.

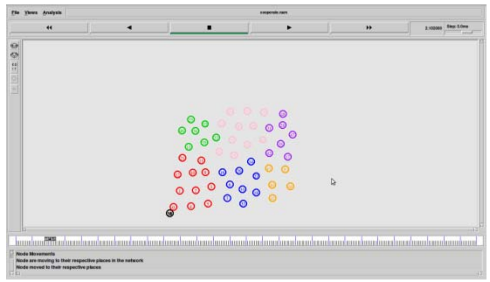


Fig 2: Nodes deployment and grid formation

5.2 Deployment of Examiner Nodes

Then we will deploy examiner nodes into the network. Each grid will have one examiner node. Examiner nodes do not participate in routing and do not forward any data packets. So, examiner nodes are not a part of the network.

Fig 3: Deployment of examiner nodes in each grid

5.3 Computing Threshold value of number of neighbors

Each node will calculate its neighbor count and will inform to the examiner node, of their corresponding region, about it. Then examiner nodes will compute the threshold value of number of neighbors of each node.

5.4 Detection of Attacker nodes

Source will broadcast the route request message to one of its neighbors. Examiner node will check if any node sending more packets than the threshold value then compares its Packet Delivery Ratio with its neighbor nodes. If any node flooding into the network, PDR of that particular node will be very high and If PDR is abnormal, examiner node will mark that node as malicious and the network will stop communicating with the node sending more packets than the threshold value.

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 3, March 2016

6. SIMULATION RESULTS

Performance of the proposed protocol early detection of DDoS attack in WSN has been examined by comparing the proposed protocol with the existing PPS (profile based protection scheme) for the detection of DDoS attack on WSN on the basis of flood count, packet delivery ratio, throughput and energy reserved. The whole scenario is tested with different number of attackers that varies from one to four.

6.1 Packet Delivery Ratio

It is defined as the ratio of no. of packets received to the no. of packets sent in the network to the base station. The greater value of the packet delivery ratio means better performance of protocol. The proposed protocol early detection of DDoS attack in WSN has the greater value of the packet delivery ratio hence have better performance in comparison of PPS.

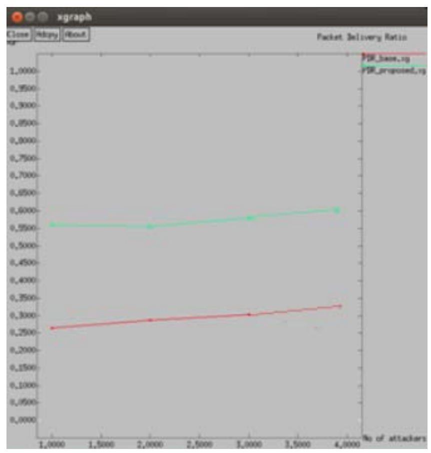


Fig 4: Comparison of PDR

TABLE 2

No. of attackers	PPS scheme	Proposed scheme
1 attacker	0.265	0.560
2 attackers	0.288	0.555
3 attackers	0.3036	0.565
4 attackers	0.31	0.5807

6.2 THROUGHPUT

Throughput is the average of data packets received at the destination. The proposed protocol early detection of DDoS attack in WSN shows the improved throughput value as compared to PPS. The more packet delivery ratio provides the improved throughput value in the network.

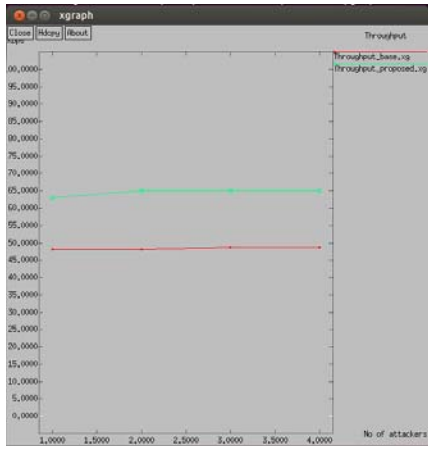


Fig 5: Comparison of throughput

TABLE 3

No. of attackers	PPS scheme	Proposed scheme
1 attacker	48	63
2 attackers	48	65
3 attackers	48.6	65
4 attackers	48.7	65

6.3 FLOOD COUNT

Flood count is number of packets flooded into the network by different number of attackers. The proposed protocol early detection of DDoS attack in WSN floods lesser number of packets as compared to PPS. Less flooding means lesser loss of data and it will help to save more energy.

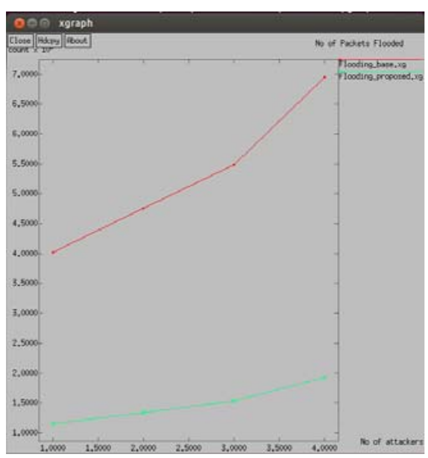


Fig 6: Comparison of flood count

TABLE 4

No. of attackers	PPS scheme	Proposed scheme
1 attacker	4021	1150
2 attackers	4754	1344

3 attackers	5486	1538
4 attackers	6950	1926

6.4 REMAINING ENERGY

The energy consumption is the aggregate of used energy by all the nodes in the network, where the used energy of a node is the sum of the energy used for transmission, including sending, receiving, and idling. In comparison of PPS scheme and the the proposed scheme, the remaining energy of the proposed protocol Early detection of DDoS attack in WSN is more because of less flooding in the network. The more remaining energy provides the more stability period and network lifetime.

Energy {exp $ initial energy ($i)-$ final energy ($f)}

Fig 7: Comparison of Remaining Energy

TABLE 5

No. of attackers	PPS scheme	Proposed scheme
1 attacker	83.6567	84.9701
2 attackers	83.5337	84.85
3 attackers	82.6566	84.7946
4 attackers	82.6563	84.2156

7. CONCLUSION

In WSN the nodes are continuously interchanging the information in network. But the information is in the form of large number of packets flooded in network then the network is assumed to be affected from DDoS attack. The proposed scheme, detect the attacker on early stages before it is completely launched into the network that prevents the data loss in the network, reserves more energy after the effect of attack, reduces the flood count in the network and enhance the throughput. The proposed scheme has been compared with the existing PPS scheme against DDoS attack in WSN and the proposed scheme is giving better results than the PPS.

8. ACKNOWLEDGEMENT

The authors wish to thank the faculty from the computer science department at CTIEMT, Jalandhar for their continued support and feedback.

9. REFERECES

[1] Varsha Nigam, Saurabh Jain and Dr. Kavita Burse, "Profile based Scheme against DDoS Attack in WSN", 2014 Fourth International Conference on Communication Systems and Network Technologies, IEEE, 2014.

[2] Shikha Jindal and Raman Maini "An Efficient Technique for Detection of Flooding and Jamming Attacks in Wireless Sensor Networks", International journal of computer applications, 0975-8887, Vol. 98, No. 10, 2014.

[3] Upavi .E.Vijay1, Nikhil Sameul2, "Study of Various Kinds of Attacks and Prevention Measures in WSN", International Journal of Advanced Research Trends in Engineering and Technology (IJARTET),Vol. II, Special Issue X, March 2015.

[4] Ahmad Abed, Alhameed Alkhatib, and Gurvinder Singh Baicher "Wireless Sensor Network Architecture," International Conference on Computer Networks and Communication Systems, IPCSIT, vol. 35, 2012.

[5] Guangjie Han, Jinfang Jiang, Wen Shen, Lei Shu, Joel Rodrigues, "IDSEP: a novel intrusion detection scheme based on energy prediction in cluster-based wireless sensor networks", IET Inf. Secur., Vol. 7, Iss. 2, pp. 97–105, , 2013.

[6] Sonali Swetapadma Sahu et.a. "Distributed Denial of Service Attacks: A Review", I.J. Modern Education and Computer Science, 2014, 1, 65-71 Published Online January 2014 in MECS.

[7] Y.-C. Hu, A. Perrig, D.B. Johnson: Adriane: A Secure On-Demand Routing Protocol for Ad Hoc Networks, Annual ACM Int. Conference on Mobile Computing and Networking (MobiCom) 2002.

[8] K. Kifayat, M. Merabti, Q. Shi, D. Llewellyn-Jones: Group-based secure communication for large scale wireless sensor networks, J. Information Assurance Security. Vol 2, 139–147, 2007.

[9] Najma Farooq1, Irwa Zahoor2, Sandip Mandal3 and Taabish Gulzar4, "Systematic Analysis of DoS Attacks in Wireless Sensor Networks with Wormhole Injection", International Journal of Information and Computation Technology,Volume 4, Number 2 (2014), pp. 173-182.

[10] Ms. Shagun Chaudhary1, Mr. Prashant Thanvi2, "Performance Analysis of Modified AODV Protocol in Context of Denial of Service (Dos) Attack in Wireless Sensor Networks", International Journal of Engineering Research and General Science Volume 3, Issue 3, May-June, 2015.

[11] Kanchan kaushal, Varsha sahni, "Early detection of DDoS attack in WSN", International Journal of Computer Applications (0975 – 8887) Volume 134 – No.13, 2016.

[12] Saman Taghavi Zargar et.al. , "A Survey of Defense Mechanisms Against Distributed Denial of Service (DDoS) Flooding Attacks", IEEE COMMUNICATIONS SURVEYS & TUTORIALS, published online Feb. 2013.

[13] Guangjie Han, Wen Shen, Trung Q. Duong, Mohsen Guizani and Takahiro Hara , "A proposed security scheme against Denial of Service attacks in cluster-based wireless sensor networks", Security and Communication Networks,Published on 9 SEP 2011.

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 3, March 2016

Detection of Stealthy Denial of Service (S-DoS) Attacks in Wireless Sensor Networks

Ram Pradheep Manohar[1] E.Baburaj[2]
[1]Research Scholar, St.Peter's University, Chennai,

[2] Professor, Narayanaguru College of Engineering, Nagercoil.

Abstract—Wireless sensor networks (WSNs) supports and involving various security applications like industrial automation, medical monitoring, homeland security and a variety of military applications. More researches highlight the need of better security for these networks. The new networking protocols account the limited resources available in WSN platforms, but they must tailor security mechanisms to such resource constraints. The existing denial of service (DoS) attacks aims as service denial to targeted legitimate node(s). In particular, this paper address the stealthy denial-of-service (S-DoS)attack, which targets at minimizing their visibility, and at the same time, they can be as harmful as other attacks in resource usage of the wireless sensor networks. The impacts of Stealthy Denial of Service (S-DoS) attacks involve not only the denial of the service, but also the resource maintenance costs in terms of resource usage. Specifically, the longer the detection latency is, the higher the costs to be incurred. Therefore, a particular attention has to be paid for stealthy DoS attacks in WSN. In this paper, we propose a new attack strategy namely Slowly Increasing and Decreasing under Constraint DoS Attack Strategy (SIDCAS) that leverage the application vulnerabilities, in order to degrade the performance of the base station in WSN. Finally we analyses the characteristics of the S-DoS attack against the existing Intrusion Detection System (IDS) running in the base station.

Index Terms— *resource constraints, denial-of-service attack, Intrusion Detection System*

I Introduction

Wireless sensor network (WSN) is a fast growing technology that is currently attracting considerable research interest. Recent advances in this field have enabled the development of low-cost, low-power and multi-functional sensors in wireless communications and electronics that are small in size and communicate in short distances. Cheap and smart sensors are networked through wireless links and deployed in large number, provide extraordinary opportunities for monitoring and controlling homes, cities, and the environment. Moreover, the sensor network has a wide range of applications in the area of defense, surveillance, generating new capabilities for reconnaissance and also for other tactical applications.

The threats in the WSN can be from outside the network and within the network. The attack in the WSN is much harmful if it is from the native network and also it is difficult to detect the malicious or compromised node within the network. The classification of the attack can be of two types: active attack and passive attack. The passive attacks do not alter or modify the data whereas the active attacks do.

The classification of the WSN attack can be done in two broad categories: invasive and non-invasive. The targets of the non-invasive attacks are timings, power and frequency of channel whereas the targets of the invasive attacks are the availability of service, transit of information, routing etc. In Denial of Service (DoS) attack tries to make system or service inaccessible. However during the transmission of information, more common attacks are also encountered. Routing attacks are generally inside attacks that occur within the network.

DoS and Distributed DoS (DDoS) aim at reducing the service availability and performance by exhausting the resources of the base station (service's host system) [1]. Such attacks have special effects in the WSN. The delay of the service to diagnose the causes of the degradation in the service (i.e., if it is due to either an attack or an overload) can be considered as a vulnerability to the security. It can be oppressed by attackers that aim at exhausting the base station resources, and seriously degrading the Quality of Service (QoS).

There are varieties of conditions for the DOS attack and these conditions may annoy the WSN nodes and network functionality. These conditions leads to the resource exhaustion, any software bug, or any other complication will be created in the application during the interaction, infrastructure and hence the normal routines of the network is disturbed. These conditions that hinder the network functionality are called as the DoS as it affects the availability or entire functionality of service but when it is caused intentionally by the opponent then it is called DoS attacks.

Many techniques have been proposed for the detection of DDoS attacks in distributed environment. Security prevention mechanisms usually use approaches based on rate-controlling, time-window, worst-case threshold, and pattern-matching methods to discriminate between the nominal system operation and malicious behaviors [2]. But the attackers are aware of the presence of such protection mechanisms. Hence the attackers attempt to perform their activities in a stealthy manner in order to escape from the security mechanisms, by planning and coordinating the attack. The timing attack patterns leverage specific weaknesses of target systems [3]. They are carried out by directing flows of legitimate service requests against a specific base station at such a low-rate that would hinder the DDoS detection mechanisms, and elongate the attack latency, i.e., the amount of time that the intruder attacking the system has been undetected.

The proposed attack strategy, namely Slowly Increasing and Decreasing under Constraint DoS Attack Strategy (SIDCAS) leverage the application vulnerabilities, in order to degrade the performance of the base station in WSN. The term under constraint is inspired to attacks which change message sequence at every successive infection in detection mechanisms [9] by using inter arrival rate of the message. Even if the victim detects the SIDCAS attack, the attack strategy can be re-initiate by using a different volume of message sequence.

The rest of the paper is organized as follows. The related work is presented in section 2. The section 3 explains in detail about the stealthy attack model. The detail about the attack approach is presented in the section 4. The evaluation of the proposed stealthy attack method is done in the section 5. Conclusion is described in section 6.

II Related work

Sophisticated DDoS attacks are defined as the attacks, which are adapted to the target system, in order to carry out denial of service or just to significantly degrade the performance of the target system [5], [8]. The term stealthy has been used in [9] to identify sophisticated attacks that are purposely designed to keep the malicious behaviors almost invisible to the detection mechanisms. These attacks can be significantly harder to detect compared with the brute-force and flooding style attacks [3].

DoS attacks can seriously degrade the network performance by interrupt the routing mechanism and thus exhausting network resources. The network layer DoS attacks in WSN can be of different category. Blackhole attack in which the malicious or compromised node absorb all the traffic going toward the target node [10], Greyhole attack in which the compromised node forwards the packets selectively to the destination node, Wormhole attack to produce routing disruptions [11], Flooding attack in which the compromised node in order to congest the network transmit the flood of packets to the target node to degrade the networks performance. A flooding DoS attacks are difficult to handle and hence an active cache based defense against the flooding style of DoS attacks is proposed in [12]; however this mechanism does not effectively handle the Distributed DoS attack. All these DoS attacks are observed in WSN due to its multi-hop nature. A distributed flooding DoS attack is a huge challenge for all the wireless sensor networks because this type of attack greatly reduces the performance of the network by consuming the network bandwidth to the large extent. This kind of denial of service attack is first launched by compromising large number of innocent nodes in the wireless network termed as Zombies [13], which are programmed by highly trained programmer. These zombies send data to selected attack targets such that the aggregate traffic congests the network. In most of the cases, the DDoS is difficult to prevent and it has the ability to flood and overflow the network [16]. In recent years, variants of DoS attacks that use low-rate traffic have been proposed some of them are Reduction of Quality attacks (RoQ), Shrew attacks (LDoS), and Low-Rate DoS attacks against application servers (LoRDAS).

Therefore, several works have proposed techniques to detect the different forms of the above mentioned denial of service attacks, which monitor anomalies in the fluctuation of the incoming traffic through either a time or frequency-domain analysis [14], [15], [16]. They assume that, the main anomaly can be incurred during a low-rate attack is that, the incoming service requests fluctuate in a more extreme manner during an attack. The two different types of behaviors are combined together to form the abnormal fluctuation: (i) a periodic and impulse trend in the attack pattern, and (ii) the fast decline in the incoming traffic volume (the legitimate requests are continually discarded).

To the best of our knowledge, none of the works proposed in the literature focus on stealthy attacks against application that run in the WSN.

III Stealthy attack model

III.1. Base Station under Attack Model

We suppose that the system consists of set of sensor nodes as clients or users and set of services provided by the Base Station (BS), on the basis of which application instances run. Moreover, we assume that a load balancing mechanism dispatches the user service requests among the instances. Specifically, we model the system under attack with a comprehensive capabilityzM, which represents a global amount of work the system is able to perform in order to process the service requests.

Fig. 1. Base Station Queue Capacity.

Such capability is affected by several parameters, such as the number of process assigned to the application, the base station performance, the memory capability, etc. Each service request consumes a certain amount of the capability zM on the base of the payload of the service request. The BS Queue Capacity is shown in the Fig. 1. The parameter 0 – no queue, zM – manageable queue and Max – maximum queue capacity (bottle -neck).

III.2. Stealthy Attack Objectives

We define the characteristics that a DDoS attack against an application running in the wireless sensor network should have to be stealthy. Regarding the quality of service of the system, we assume that the system performance under a DDoS attack is more degraded, as higher the average time to process the user service requests compared to the normal operation.

The stealthy attackers aim is that a complicated attacker would like to achieve, and the requirements the attack pattern has to satisfy to be stealth. The purpose of the attack against wireless sensor applications is not to necessarily deny the service, but rather to impose significant degradation in some aspect of the service (e.g., service response time), namely benefit of attack BA, in order to maximize the base station computation cost CC to process malicious requests. Therefore, in order to perform the attack in stealthy fashion with respect to the proposed detection techniques, an attacker has to inject low-rate message flows;

$$MF = \{mf(A_{j,1}), mf(A_{j,2}), \ldots, mf(A_{j,m})\} \tag{1}$$

where $j = 1, 2, \ldots, N$ is the number of Attackers and $m = 1, 2, \ldots, M$ is the number of messages. Stealthy DoS attack pattern in WSN denote p the number of attack flows, and consider a time window T, the DoS attack is successful in the WSN, if it maximizes the following functions of Benefit of Attack (BA) and Computation Cost (CC):

$$Maximize\ BA = \sum_j \sum_m B[mf(A_{j,m})] \tag{2}$$

where B is the benifit of the malicious request $A_{j,m}$, which expresses the service degradation (e.g., in terms of increment of average service time t_S to process the user requests with respect to the normal operation);

$$Maximize\ CC = \sum_j \sum_m W[mf(A_{j,m})] \tag{3}$$

where W is the computation cost in terms of base station resources necessary to process.

III.3. Creating Service Degradation

Considering a base station with a comprehensive capability zM to process service requests $mf(N_j)$, and a queue with size Q that represents the bottleneck shared by the customer's flows $mf(N_j)$ and the DoS flows $mf(A_j)$. That is the base station work under the safe condition (not in bottleneck stage) under the condition;

$$T\left(\sum_j mf(N_j) + \sum_j mf(A_j)\right) < zM \tag{4}$$

where $mf(N_j)$ is the normal nodes message flows, $mf(A_j)$ is the attacker nodes message flow and zM is the base station safe stage threshold. So that, number of message flows in time $T\left(\sum_j mf(N_j) + \sum_j mf(A_j)\right) > zM$ the base station under service degradation stage.

III.4. Minimize Attack Visibility

According to the stealthy attack definition, in order to reduce the attack visibility the attacker exhibits a pattern neither periodic nor impulsive and also exhibits a slowly increasing intensity in the attack rate. Therefore, through the analysis of both the attacker system and the normal service requests not exceed the base station safe stage threshold zM. So that the attacker system maintains the stealthy attack by balancing the message flows.

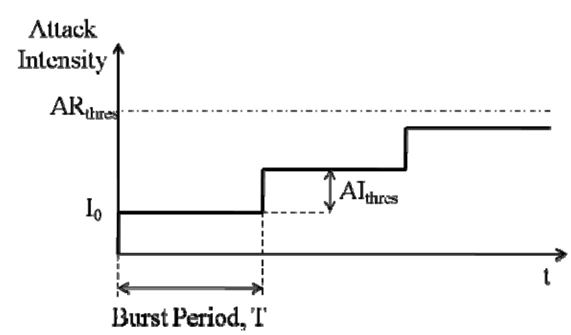


Fig. 2. Increment of stealthy attack intensity.

To implement an attack pattern that maximizes BA and CC, as well as satisfies stealthy condition, without knowing in advance the target system characteristics, we propose a attack strategy, which is an iterative and incremental process. At the first iteration only a limited number p of flows $mf(A_j)$ are injected. The value p is increased by one unit at each iteration p, until the desired service degradation is achieved.

During each iteration, the flows $mf(A_j)$ exhibit the attack intensity shown in Fig. 2. Specifically, each flow $mf(A_j)$ consists of burst of messages, in which the parameter $I_0(p)$ means the initial attack intensity at the iteration p (which can be orchestrated by varying the number and type of injected requests), T is the length of the burst period, and AI_{thres} is the increment of the attack intensity each time a specific condition AR_{thres} is false. AR_{thres} is tested at the end of each period T. The satisfaction of the condition AR_{thres} identifies the achievement of the desired service degradation.

IV Attack approach

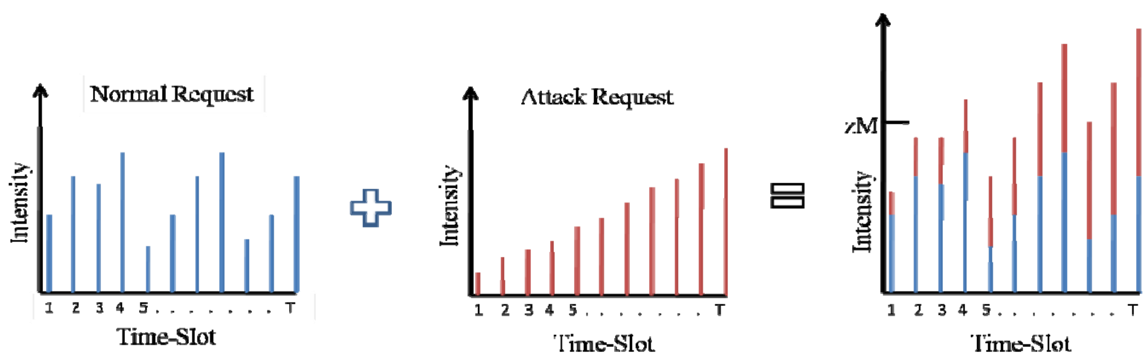
In order to implement SIDCAS-based attacks, the following components are involved:

- a Master that coordinates the attack A^a;
- p Agents that perform the attack A^p, each Agent injects a single flow of messages $mf(A_j)$; and
- A^a Meter that evaluates the attack effects.

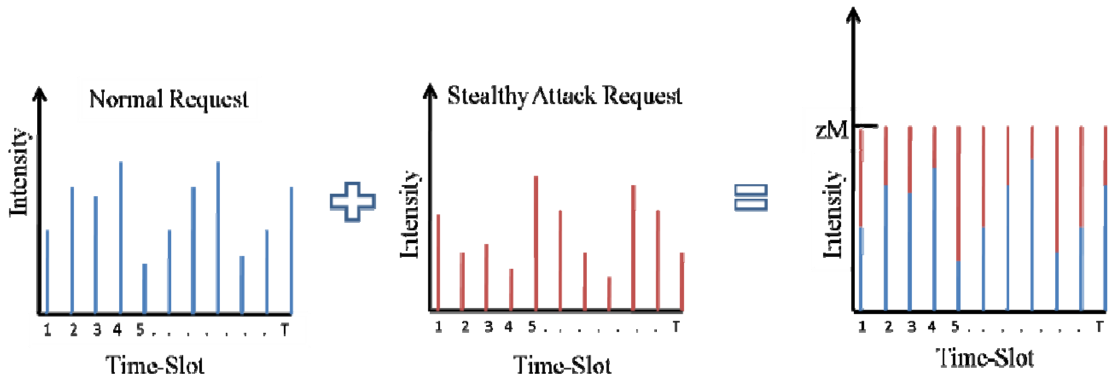
Algorithm 1 describes the approach implemented by each Agent to perform stealthy service degradation in the WSN. Specifically, the attack is performed by injecting polymorphic bursts of length T with an increasing intensity until the attack is either successful or detected and t_I is the inter-arrival time between two consecutive requests. Each burst is formatted in such a way as to inflict a certain average level of load C_R. In particular, we assume that C_R is proportional to the attack intensity of the flow $mf(A_j)$ during the period T. Therefore, denote I_0 as the initial intensity of the attack.

Algorithm 1: Working Algorithm of SIDCAS-based Attack

Require: $TimeWindow\ T$
Require: $Attackratethreshold\ AR_{thres}$
Require: $Attackintensityincrement\ AI_{thres}$
Require: $Initialattackintensity\ I_0$
$t \Leftarrow 0;$
while $t \leq T$ **do**
    $t_I \Leftarrow computeInterarrivalTime(C_R);$
    $sendMessage(t_I);$
    $t \Leftarrow t + t_I;$
end while
if $!\,(attackSuccessful)$ **then**
    $C_R \Leftarrow (C_R + attackIncrement);$
    $\{Attack\ intensification\}$
else
    while$!\,(attack\_detected)$ **and** $attackSuccessful$ **do**
        $\{Service\ degradation\ achieved; attack\ intensity\ is\ fixed\}$
        $t_I \Leftarrow compureinterarraivalTime(C_R);$
        $sendmessage(t_I);$
    end while
end if



(a)



(b)

Fig. 3. Resultant Attack Strategy (a) Existing methods (b) Proposed Stealthy Attack.

The attack intensity in case of the normal attack strategy and the stealthy attack strategy is shown in the Fig.3. In the case of the normal existing attack strategy in Fig. 3(a), the attack intensity increases linearly towards a high value and hence the intensity of the message request increases beyond the maximum intensity of the queue. This dramatic increase in the request intensity allows the server to detect the presence of the attacker and prevention measures will be taken by the server. But in the case of the stealthy attack pattern in Fig. 3(b) the attack intensity increases iteratively and incrementally. Also the attack intensity does not exceed after the maximum attack intensity and hence the server will not be able to determine the presence of the attacker.

V. Performance evaluation

The effectiveness of the proposed stealthy attack can be evaluated with the Attack Detection Ratio (ADR) and the Resource Usability (RU). The ADR is the detection rate of the attacker request by the base station and it is given by the Equation (4):

$$ADR = \frac{No.of\ Attacker\ Detected}{Total\ No.of\ Attacker} \tag{4}$$

The ADR of the DOS attack and the Stealthy attack is shown in the Fig. 4. From the comparison plot it can be seen that the detection rate of the DOS attack increases as the number of the attacker increases but the ADR value remains lower for the stealthy attack pattern even if the number of attacker increases.

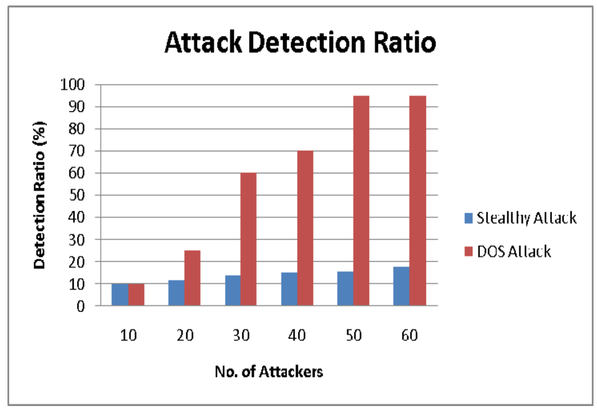


Fig. 4. Attack Detection Ratio.

The stealthy attack pattern mainly concentrates on the resource usability rather than the denial of service. Hence the RU of the base station in case of the DOS attack and the proposed stealthy attack is shown in the Fig. 5.

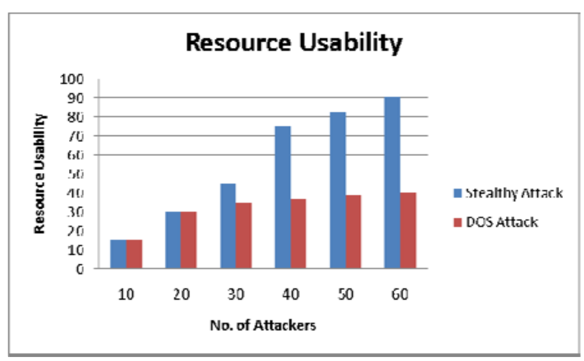


Fig. 5. Resource Usability.

From the plot it can be shown clearly that the stealthy attack pattern utilizes more resources as the number of attacker increases. Because as the number of attacker is increased in the stealthy manner the base station cannot detect the presence of the attacker and this will lead to more resource usability even if the number of authorized nodes in the queue is low.

The RU is also dependent on the time required for the base station for processing the request. That is, in the case of the DOS attack if the number of attacker increases the presence of the attacker will be detected by the IDS and hence the processing time required for the attack request will be reduced. But in the case of the stealthy attack pattern the even if the number of attacker is more the presence of the attacker will not be detected by the IDS and hence more resources will be utilized for the processing of the attacker request.

VI. Conclusion

In this paper, we propose a new strategy to implement stealthy attack patterns in WSN, which reveal a stealthy behavior that can be greatly unrecognizable by the techniques proposed in the existing intrusion detection system against the DoS attacks. For developing a vulnerability of the target base station or access point in the WSN, an intelligent attacker can organize a customize or dynamic flows of access, indistinguishable from legitimate access requests. In particular, the proposed attack pattern, instead of aiming at making the access unavailable, it aims at make use of the resources, forcing the system to consume more resources than needed, affecting the entire network more on resource aspects than on the access availability. In the future work, we aim at developing an approach that able to detect stealthy nature attacks in the wireless sensor network environment.

References

[1] K. Lu, D. Wu, J. Fan, S. Todorovic, and A. Nucci, “Robust and efficient detection of DDoS attacks for large-scale internet,” Comput. Netw., vol. 51, no. 18, pp. 5036–5056, 2007.

[2] H. Sun, J. C. S. Lui, and D. K. Yau, “Defending against low-rate TCP attacks: Dynamic detection and protection,” in Proc. 12th IEEE Int. Conf. Netw. Protocol., 2004, pp. 196-205.

[3] A. Kuzmanovic and E. W. Knightly, “Low-rate TCP-Targeted denial of service attacks: The shrew vs. the mice and elephants,” in Proc. Int. Conf. Appl., Technol., Archit., Protocols Comput. Commun., 2003, pp. 75–86.

[4] M. Guirguis, A. Bestavros, I. Matta, and Y. Zhang, “Reduction of quality (RoQ) attacks on internet end-systems,” in Proc. IEEE Int. Conf. Comput. Commun., Mar. 2005, pp. 1362–1372.

[5] X. Xu, X. Guo, and S. Zhu, “A queuing analysis for low-rate DoS attacks against application servers,” in Proc. IEEE Int. Conf. Wireless Commun., Netw. Inf. Security, 2010, pp. 500–504.

[6] L. Wang, Z. Li, Y. Chen, Z. Fu, and X. Li, “Thwarting zero-day polymorphic worms with network-level length-based signature generation,” IEEE/ACM Trans. Netw., vol. 18, no. 1, pp. 53–66, Feb. 2010.

[7] U. Ben-Porat, A. Bremler-Barr, and H. Levy, “Evaluating the vulnerability of network mechanisms to sophisticated DDoS attacks,” in Proc. IEEE Int. Conf. Comput. Commun., 2008, pp. 2297–2305.

[8] S. Antonatos, M. Locasto, S. Sidiroglou, A. D. Keromytis, and E. Markatos, “Defending against next generation through network/ endpoint collaboration and interaction,” in Proc. IEEE 3rd Eur. Int. Conf. Comput. Netw. Defense, 2008, vol. 30, pp. 131–141.

[9] H-.M. Deng, W. Li, D.P. Agarwal, “Routing Security in Wireless Ad Hoc Networks,” IEEE Communication Magazine, Vol. 40, pp. 70-75, October 2002.

[10] F.N. Abdesselam, B. Bensaou and T. Taleb, “Detecting and avoiding wormhole attacks in wireless Ad hoc networks,” IEEE Communication Magazine, Vol.46, Issue 4, pp. 127-133, April 2008.

[11] L. Santhanam, D. Nandiraju, N. Nandiraju and D. P. Agrawal, “Active cache based defence against DoS attacks in Wireless Mesh Network,” Proceedings of the 2nd IEEE Int. Symp.Wireless Pervasive Computing (ISWPC2007), 2007.

[12] G.A Marin “Network security basics,” In IEEE Security and Privacy, Vol.3, p 68-72, November 2005.

[13] A. Kuzmanovic and E. W. Knightly, “Low-rate TCP-targeted denial of service attacks and counter strategies,” IEEE/ACM Trans. Netw., vol. 14, no. 4, pp. 683–696, Aug. 2006.

[14] X. Luo and R. K. Chang, “On a new class of pulsing denial-of-service attacks and the defense,” in Proc. Netw. Distrib. Syst. Security Symp., Feb. 2005, pp. 61–79.

[15] Y. Chen and K. Hwang, “Collaborative detection and filtering of shrew DDoS attacks using spectral analysis,” J. Parallel Distrib. Comput., vol. 66, no. 9, pp. 1137–1151, Sep. 2006.

[16] P. Dasgupta, T. Boyd, “Wireless Network Security,” Annual Review of Communications, Vol.57, International Engineering Consortium, 2004.

Intelligent Radios in the Sea

Ebin K Thomas
Amrita Center for Wireless Networks & Applications (AmritaWNA)
Amrita School of Engineering, Amritapuri
Amrita Vishwa Vidyapeetham
Amrita University
India

Abstract- **Communication over the sea has huge importance due to fishing and worldwide trade transportation. Current communication systems around the world are either expensive or use dedicated spectrum, which lead to crowded spectrum usage and eventually low data rates. On the other hand, unused frequency bands of varying bandwidths within the licensed spectrum have led to the development of new radios termed Cognitive radios that can intelligently capture the unused bands opportunistically by sensing the spectrum. In a maritime network where data of different bandwidths need to be sent, such radios could be used for adapting to different data rates. However, there is not much research conducted in implementing cognitive radios to maritime environments. This exploratory article introduces the concept of cognitive radio, the maritime environment, its requirements and surveys, and some of the existing cognitive radio systems applied to maritime environments.**

Keywords— Cognitive Radio, Maritime Network, Spectrum Sensing.

I. INTRODUCTION

Current technology growth is mind blowing. Terrestrial or land based communication systems have seen tremendous growth in the form of 3G, 4G, WiMAX, LTE and LTE Advanced but, this growth has not reflected maritime networks since most of the marine communication systems are primitive or in under-developed state [1].

A maritime communication system is essentially a network architecture comprising of equipments such as base stations, clients (ships and boats) and user end devices capable of communicating over a sea environment. Such an environment is entirely different from land in terms of the atmospheric conditions, wireless channel properties that affect coverage ranges. Hence, existing terrestrial communication systems cannot be directly applied to marine environments. Communication over the sea can either be Line of Sight (LOS) or Non-Line of Sight (NLOS). LOS communication means a direct path between a transmitter and receiver up to a certain distance,whereas NLOS communication occur due to obstructions. Obstructions in terrestrial communications include trees and buildings in contrast to sea where the only obstruction is the Earth's horizon or the bulge. Figure.1 below depicts a transmitter and a receiver at LOS across the sea with the Earth's bulging effect.

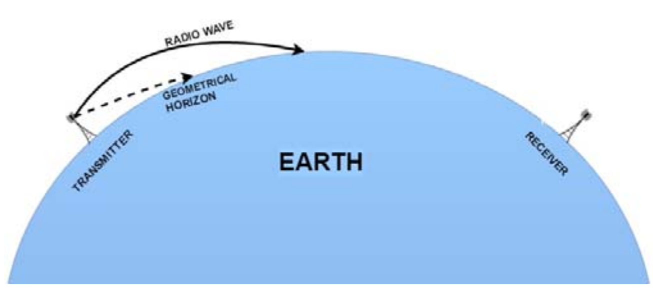


Figure.1 Earth horizon

LOS communication in a marine environment is made possible through radio or microwave frequencies. Even though, there is a dedicated spectrum for maritime communication, it is limited in bandwidth and thus data rates are very low. Therefore, there is a need for alternative methods to enable communication at sea, which provides improved data rates, coverage, and connectivity. Table below shows the typical maritime frequency bands and their theoretical data rates.

Table-I

System	Frequency	Data rate
VHF voice	156.300 MHz, 156.650 MHz, 156.800 MHz	25 kHz
AIS	161.975 MHz (AIS 1) 162.025 MHz (AIS 2)	9.6 kbps
Satellite	1626.5 to 1646.5 MHz 1525.0 to 1545.0 MHz	256kbps

From Table I, communication systems based on UHF and VHF band are used for ship-to-shore and ship-to-ship communication but are limited in capacity. Satellite communication is preferred for long range, but the trade-offs include cost and infrastructure set up. A typical satellite system costs around Rs. 25,000 (USD 500). All these systems make use of a dedicated spectrum, which is difficult due to overcrowding of these bands, eventually leading to congestion. Moreover, the network devices on the shore may need to coexist with other radio devices installed on the land and they also need to synchronize the frequency bands around the world when the ships move across different countries and continents [10]. Another main issue is the ineffective use of spectrum leading to spectrum gaps [3]. Figure. 2 shows the utilization of the spectrum.

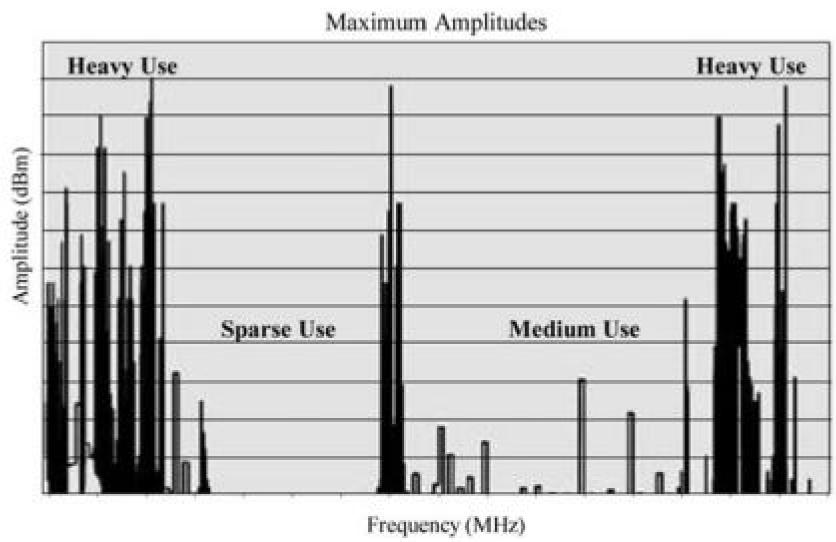


Figure.2: Spectrum utilization [3]

Figure.2 depicts that the frequency bands are not used efficiently, leaving *white spaces*. Existing hardware based radios are incapable of being tuned to different frequency bands, leaving the unused bands wasted. Cognitive radio is the key enabling technology that enables next generation communication networks to utilize the spectrum more efficiently in an opportunistic way. This article is intended to provide the readers an overview of cognitive radio and its application in the maritime environment. We discuss the working of a cognitive radio, the parameters and the factors that affect the functioning of a cognitive radio, the maritime environment, the challenges it faces and finally review some existing works in this domain. To summarize, our contributions are:

1. Provide a succinct tutorial style review of cognitive radios
2. Open new frontiers in cognitive radio research and its applications
3. Propose a light weight cognitive radio network architecture for a maritime environment

The paper consists of the following sections. Section II discusses the concept of cognitive radio and its features. Section III introduces the maritime environment. Section IV reviews the existing networks proposed for maritime environments. Section V provides a discussion on the need for a cognitive radio in a maritime environment, followed by a simulation set up and finally Section VI concludes the paper.

II. COGNTIVE RADIO

What is a cognitive radio?
A cognitive radio is a radio with intelligence, capable of identifying neighbouring available free channels or spectrum , that can be accessed and used for transmission. It was first invented and coined by Joseph Mitola [2].

Need for a cognitive radio:
Inefficient use of the spectrum can lead to shortage of channels necessary for transmission. In-order to solve this problem, the concept of cognitive radio (CR) was proposed. Here the radio is programmed to *sense*, *acquire* and *utilize* the spectrum bands available for a certain period of time, thereby mitigating spectrum scarcity. This intelligent use of spectrum is termed Dynamic Spectrum Access (DSA) [4].

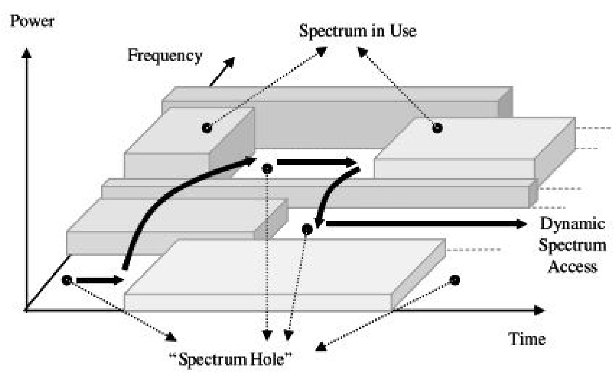


Figure 3: White space or spectrum gaps [3]

How does a cognitive radio work?
A typical cognitive radio cycle consists mainly of five steps as shown in Figure. 4.

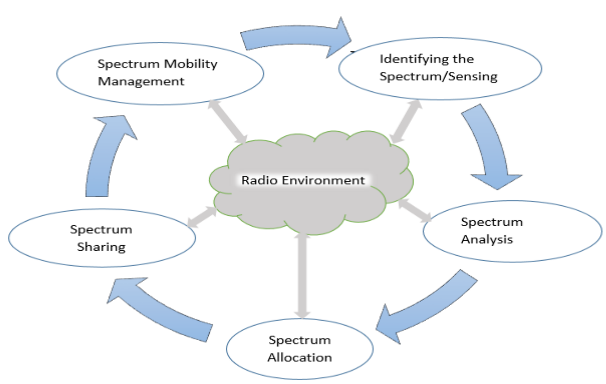


Figure 4: The cognitive radio cycle

First the spectrum is identified(sensed), then detected, if the spectrum is being used (primary user detection). This is followed by an analysis of the spectrum characteristics and allocating it for use. In some cases, the spectrum can also be shared among multiple secondary users which leads to proper spectrum management.

A. Spectrum Sensing

A white space (WS) is defined as an unused or a currently inactive frequency band (i.e., when not being used by primary users (PU)).Therefore, intelligent radios or secondary users (SU) need to look at how to identify such WSs efficiently without creating any interference to the PUs. Methods to identify the available spectrum include beacon based method, geolocation database, mechanisms which detect various signal features, energy of the PU's signal, matched filter and cyclo-stationary detection. We briefly discuss these mechanisms below.

a. Beacon based method[4]

A beacon is deployed to send signals indicating the presence or absence of PU. The secondary device can transmit only after it receives some signal. The major drawback with this method is the need for additional circuitry on the transmitter side since it adds some extra cost and the secondary device cannot transmit the data whenever it wants.

b. Geolocation data base [4]

In geolocation database, a database is created using the information about all available WSs in the TV transmission bands at different locations. Whenever a secondary device needs to access the spectrum, it has to look up the database and transmit the data in any available band. Without proper interference reduction technique, simply looking up and accessing the WS can cause interference to the PU. Moreover, creating a database is complex, requiring continuous scanning, consistent updates,which all add to the computational cost on the hardware.

c. Feature Detection

Every primary signal has features such as modulation rate, carrier frequency, wave pattern, signal power and so on, that are extracted and compared with the sensed signal to identify the presence of the PU.The popular feature detection mechanisms are explained below.

i. Matched Filtering and Coherent Detection

In this method ,the secondary user knows about PU's wave pattern and the sensed signal will correlate with primary user signal to detect the presence of the primary signal. This method can also distinguish between noise signals even if the sensed signal amplitude is low.

ii. Energy Detection

In this method, SU will sense for the presence of the PU through the PU's signal power. Signal power is defined as the amount of energy consumed over a unit time. If the signal originating from the PU is greater than a threshold, the SU concludes a PU being present. However, this method is not robust when the Signal to Noise ratio (SNR) reaches close to the receiver sensitivity (-89dBm to -91dBm). Another problem is the SU might falsely detect the presence of a PU in scenarios where there is some noise signal which has higher energy than the threshold value.

iii. Cyclo-stationary detection

Cyclo-stationary means signal characteristics such as frequency and modulation rate are periodic and occur in a cyclic form. Therefore, these features are spectrally highly correlated in contrast to noise which is aperiodic. In cyclo-stationary detection these features are examined to distinguish between the primary and the secondary signal.

d. Entropy based spectrum sensing

Entropy is defined as the average amount of information in a signal. For a given signal power, the entropy of a signal will decrease the presence of primary signal or any modulated signal and entropy is maximized in the presence of a noise signal. The currently available best method is a combination of matched filter and entropy measurement. The advantage of this method is that it does not need any prior information about the signal or noise.

B. Spectrum Analysis

Once a freely available spectrum is detected using any of the methods discussed above, issues such as the SUs have to analyze the spectrum based on the user requirement before allocating for a particular user. The spectrum should also be analyzed for the available bandwidth and the kind of data that can be sent over this spectrum.

C. Spectrum management

After acquiring the best available spectrum based on the user requirements, the SU has to choose the best spectrum band to meet the Quality of Service (QoS) requirements. Since the spectrum requirements vary from time to time and location to location, parameters like interference, path loss, link quality, and channel capacity needs to be considered before allocating the spectrum.

D. Spectrum mobility

Spectrum mobility is caused by two events that trigger the SU to change its frequency of operation. It can be due to the reappearance of the PU or poor QoS in the current frequency band. Spectrum mobility results in the spectrum handoff, which means shift in the frequency of operation.

E. Spectrum sharing

Spectrum sharing means co-existence of SUs with licensed PU. Sometimes the PU may not use the entire spectrum, they can share the available share of spectrum in a pool known as spectrum pool. Then the secondary users can choose the spectrum from this pool. Co-existence of different SU is also important in situations where the unused spectrum is limited. In such a scenario, one of the SU has to share the spectrum with other SUs also.

This section introduced the concept of cognitive radio and its features. In the next section, we introduce a maritime environment, the challenges posed by the environment and the requirements for deploying a communication system, followed by a brief discussion on existing maritime systems proposed in the literature.

Hardware behind a Cognitive Radio:

A typical communication device have Radio Frequency (RF) front end and baseband processing unit. An ideal cognitive radio device has to work in a large frequency spectrum so this results in the modification of the existing hardware architecture, i.e., the RF front end should be capable of sensing the wide frequency band, therefore we need a wide sensing antenna which can sense over the wide frequency band. Similarly ,we need an adaptive filter and an amplifier which works with all frequency bands. The baseband processing unit also should be adaptable to the wide frequency band.

III. THE MARITIME ENVIRONMENT

What is a maritime environment?

Maritime environment is an environment pertaining to a sea which comprises of ships, boats, vessels ,etc., engaging in fishing, inter country business and security related operations. Such an environment requires a backbone network that can enable communication between ships, boats, vessels and between the land based stations. As mentioned earlier, there are few systems proposed and working in maritime communications, but not all of them are affordable or can reach the masses particularly the poor fishermen in coastal areas in India.

Challenges posed by the Maritime Environment:

Maritime environments are entirely different from terrestrial.The presence of vast water body brings in unique environmental characteristics such as reflections from the sea surface, increased humidity, sea roughness levels, and persistent rainfall. These factors tend to have a profound effect on the propagation medium alias the wireless channel. A channel is a physical transmission medium for transmitting and receiving data. The aforementioned factors cause propagation losses, i.e., attenuation of the signal as it passes through the channel and reaches the receiver. This loss in turn affects the received signal strength (RSS). Hence the signal quality degrades as it propagates through the marine environment. Some of the natural phenomenon affecting the maritime environment are summarized as follows :

Reflection

Reflection occurs when the signal travels along a surface. The sea surface is highly reflective than the land. When a signal propagates in the sea environment due to reflection, multiple signals are generated. These multipath signals degrade the actual signal due to the destructive interference.

Refraction

When a wave propagates from one medium to another due to the variation in the refractive index, the direction of the wave will change. In marine environment the refractive index varies rapidly due to the weather conditions.

Diffraction

When the wave path is obstructed by a sharp or irregular object, then wave tends to bend around the obstacle and, this is known as diffraction. The diffraction depends on the geometry of the object, phase and amplitude of the wave.

Having discussed the maritime environment and its features, we present a table that lists and compares some maritime communication systems proposed in the literature based on factors such as operating frequency, coverage distance, bandwidth and data rates are achieved.

IV. EXISTING MARITIME COMMUNICATION SYSTEMS

In this section, we highlight some of the existing maritime networks, both traditional and cognitive based and point some of the pros and cons of each network.

A. Satellite communication [7]

Satellite communication systems are widely used by the marines due to its wide coverage over the deep sea.The data rates provided by the satellite communication systems are very low and the cost per bit is high when compared with other communication systems.The cost for the equipment is also very high and it does not support multimedia applications.

B. Automatic Identification System (AIS) [7]

It operates in the VHF maritime band. It is used to identify the vessels by exchanging the data with AIS base station and other ships.

C. WISE-PORT [10]

In Singapore, WISE-PORT (Wireless-broadband-access for Seaport) provides IEEE802.16e-based wireless broadband access up to 5 Mbps, with a coverage distance of 15 km.

D. TRITON [9]

TRI-media Telematic Oceanographic Network developed for the high speed and low cost maritime communications for the vessels which are close to the shore. This project is based on IEEE 802.16d mesh technology. The authors developed a prototype that operates at 2.3 and 5.8 GHz. This system works well when there is enough number of ships , since it is using the wireless mesh technology. If the ship density goes low, then this system has an intelligent middleware which can switch to satellite system to provide the connectivity.

E. NORCOM [10]

The first digital VHF network with a data rate of 21 and 133 Kbps with a coverage range of 130 km was developed in Norway[10]. This system operates in the licensed VHF channel, which results in a narrow bandwidth and slow communication speed.

F. MICRONet [9]

A solution based on the Long Range Wi-Fi (LR Wi-Fi) technology was proposed to provide seamless connectivity to fisherman. Long distance links ranging from 17-46 km was successfully setup. The links facilitated the use of VoIP applications such as Skype and WhatsApp on the dynamic over-the-sea wireless channel in the 2.4GHz frequency band.

G. Cognitive Maritime Mesh Networks [9]

Work done in this paper included a cognitive mesh network to provide high data rate communication systems in the marine environment. A mesh network is a type of adhoc network where every node relays data to its neighbouring node. Such a network is formed by the different boats in the sea for improving the communication range by utilizing the different paths from the transmitter and receiver via mesh.

H. MCRN [10]

Maritime cognitive radio networks (MCRNs) is the proposed solution which is based on a cognitive radio technology. They are sensing the spectrum by the entropy-based detector with the optimal number of samples as a local detector and a decision on the availability of the spectrum is made with help of the cooperative spectrum sensing.

System	Frequency	Band	Range In KM	Data rate	Cost Approximately	Technology	Existing system can made as intelligent or not	System implemented / proposed
Satellite Communication	Transmitter 1626.5 to 1646.5 MHz Receiver 1525.0 to 1545.0 MHz	UHF	100-6000	256kbps	$6000	Satellite	No	Implemented
Automatic Identification System (AIS)	161.975 MHz and 162.025 MHz	VHF	22-37	2-3kbps	$2500	Maritime VHF	No	Implemented
Digital VHF in Norway	Maritime VHF	VHF	130	21 and 133kbps	Low	Maritime VHF	No	Implemented
WISE-PORT	2.4,3.5GHz	UHF/ SHF	15	5 Mbps	Low	IEEE802.16		Implemented
TRITON	5.8Ghz	UHF/ SHF	14	6Mbps	Low	IEEE802.16	Yes	Implemented
MICRONet	2.4 and 5.8	UHF/ SHF	17-46	3.2Mbps	Low	Long range Wi-Fi	Yes	Implemented
MCRN(Maritime cognitive radio networks)	TVWS	VHF/ UHF			Low	Cognitive Radio, Satellite	Yes	Proposed

V. WHY COGNITIVE RADIO IN A MARITIME ENVIRONMENT?

A cognitive or intelligent radio could be used for an on-demand data transmission. In other words, depending on the type of message that needs to be sent, the radio can sense the available frequency bands and smartly send a message based on the bandwidth available. As a part of a funded project named MICRONet [9], we envision a maritime network capable of sending and receiving various types of messages. Table III lists the type of messages, the approximate bandwidth required and some examples of such messages.

Table III: Message types, Applications and their bandwidth requirement

Type of message	Bandwidth requirement	Examples
Instant messaging	1kbps	WhatsApp,Viber
Email	50kbps	Gmail,Hotmail
VoIP	100kbps	imo
Video call	1Mbps	Skype,WhatsApp
File transfer	2Mbps	Apache Camel

From Table III, we can see that the bandwidth requirement varies according to the amount of data to transmit or receive. In an emergency situation for sending an alert message, the data may be a simple text message for which the bandwidth requirement is low. Whereas for multimedia applications such as Skype/WhatsApp or web browsing the bandwidth requirement increases. In such scenarios, trying to communicate in one frequency band may lead to a shortage or wastage of resources. Hence, a radio that is smart enough to identify the WSs and transmit the appropriate amount of data is needed.

V. CONCLUSION

Spectrum usage is becoming congested in all walks of applications. Fixed frequency usage results in congestion and unused gaps, resulting in unwanted wastage of frequency bands. There is also a dearth of an efficient maritime communication system.Therefore, a maritime network is challenging to set up because of the various challenges posed by the sea environment and hence to solve all the above challenges,we found that Cognitive radio is the most feasible technology. This article provided a simple and concise discussion on cognitive radio networks and their features. We also discussed a maritime environment, the challenges it presents and what kind of systems are already proposed in the literature and finally shed light on why a cognitive radio is required for a maritime network.

REFERENCES

[1]. Kidston, D.; Kunz, T. Challenges and opportunities in managing maritime networks. IEEE Comm. Mag, 2008, 46, 41-49.

[2]. J. Mitola and G. Q. Maguire, “Cognitive Radio Making Software Radio More Personal.” IEEE Pers. Commun. , vol. 6, no.4, Aug.1999, pp.13-18.

[3]. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, “Next generation/dynamic spectrum access/cognitive radio wireless networks: survey.” Computer Netw.. vol.50, pp. 2127-2159, May 206.

[4]. M. Nekovee, “Dynamic spectrum access—concepts and future architectures,” BT Technology Journal, vol. 24, no. 2, pp.111–116, 2006.

[5]. Nagaraj, S.V. Entropy-based spectrum sensing in cognitive radio. Signal Proc. 2009, 89, 174–180.

[6]. Pierson, W.J., Jr.; Moskowitz, L. A proposed spectral form for fully developed seas based on the similarity theory of S. A. Kitaigorodskii. J. Geosci. Res. 1964, 69, 5181–5190.

[7]. Khurram Shabih Zaidi 1 , Varun Jeoti 2 , Azlan Awan, Wireless Backhaul for Broadband Communication Over Sea. IEEE International Conference on Communications 2013.

[8]. Pathmasuntharam, J.S.; Kong, P.Y.; Zhou, M.T.; Ge, Y.; Wang, H.; Ang, C.-W.; Su, W.; Harada, H., TRITON: High Speed Maritime Mesh Networks. Proceedings of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Cannes, France,15 September 2008; pp. 1–5.

[9]. Unni, S.; Raj, D.; Sasidhar, K.; Rao, S., "Performance measurement and analysis of long range Wi-Fi network for over-the-sea communication," in 13th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), 2015, vol., no., pp.36-41, 25-29 May 2015

[10]. Waleed Ejaz, Ghalib A. Shah, Najam ul Hasan, Hyung Seok Kim., "Optimal Entropy-Based Cooperative Spectrum Sensing for Maritime Cognitive Radio Networks," ISSN 2013.

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 3, March 2016

Developing Context Ontology using Information Extraction

Ram kumar #1, Shailesh Jaloree #2, R S Thakur *3

Applied Mathematics & Computer Application, SATI
Vidisha, India

* *MANIT*
Bhopal, India

***Abstract*—Information Extraction addresses the intelligent access to document contents by automatically extracting information applicable to a given task. This paper focuses on how ontologies can be exploited to interpret the contextual document content for IE purposes. It makes use of IE systems from the point of view of IE as a knowledge-based NLP process. It reviews the dissimilar steps of NLP necessary for IE tasks: Rule-Based & Dependency Based Information Extraction, Context Assessment.**

I. INTRODUCTION

As the amount of textual information is exponentially growing, it is more than ever a key issue for knowledge management to construct an intelligent tools and methods to give access to document content and extract proper information. Information Extraction (IE) is one of the core research fields that attempt to fulfil this need. It aims at automatically extracting domain specific data from free or semi-structured contextual documents. [1]

The information extraction technique identifies key terms and relationships within the text. It performs this by finding for predefined sequences in the text, a method called pattern matching. The software infers the relationships between all the known places, people, and time to give the user with meaningful information. This technology is very useful when dealing with huge volumes of text. This paper stresses on the importance of ontological knowledge to perform each step and presents IE-based methods for the acquisition of the required knowledge as shown in Fig.1.

A. *Ontology*

Ontology, in its unique meaning, is a branch of philosophy (specifically, metaphysics) concerned with the nature of existence. It includes the identification and study of the categories of things that exist in the universe. One scenario of ontology is in Artificial Intelligence, where it is defined as "ontology is a formal, explicit specification of shared conceptualization. This definition is given by Gruber [2] which is most commonly used by knowledge engineering community. Here Conceptualization is a "world view" that is present as a set of concepts and their relations. It is the abstract representation of a real world entity (view) with the help of domain relevant concepts [3]. Since the ontologist has huge amount of knowledge which is unstructured and it should be organized. Conceptualization helps to organize and structures the acquired knowledge by use of external representations that are independent of the implementation languages and environments [4].

Now a day's ontology has been attracting a lot of attention recently since it has emerged as a very important discipline in the areas of knowledge representation [5]. Ontology refers to the shared understanding of a domain of interest and is represented by a set of domain related concepts, the relationships among the concepts, functions and instances [6].Ontology is used for representing the knowledge of a domain in a formal and machine understandable form in many areas like intelligent information processing. Thus it provides the platform for effective extraction of information and many other applications [7]. It is very useful for expressing and sharing the knowledge of semantic web.

B. *Contextual Ontology*

Contextual ontologies provide descriptions of concepts that are context-dependent, and that may be used by some user communities. Contextual ontologies do not disagree with the usual assumption that ontology provides a shared conceptualization of some subset of the real world. Many different conceptualizations may exist for the same real world phenomenon, depending on many factors. The objective of contextual ontologies is to gather into a single organized description several alternative conceptualizations that fully or partly address the same domain. The proposed solution has several advantages: First, it allows maintaining consistency among the different local representations of data. So, update propagation from one context to another is possible. Moreover, it enables navigating among contexts (i.e. going from one representation to another).

Beyond the multiple description aspect, the introduction of context into ontologies has many other important benefits. Indeed, querying global shared ontologies is not a straightforward process as users can be overburdened with lots of information. In fact, a great deal of information is not always relevant to the interest of particular users. Ontologies should provide each user with only the specific information that he or she needs and not provide every user with all information. The introduction of the notion of context in ontologies will also provide an adapted and selective access to information.

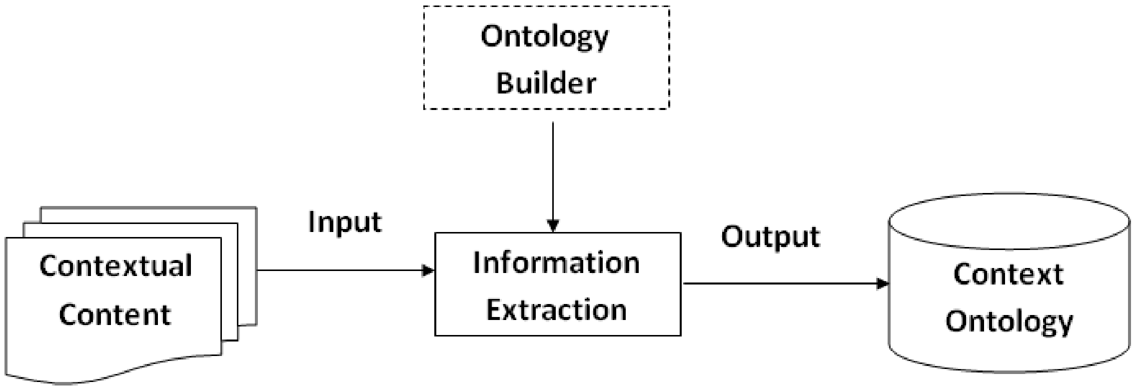


Fig.1: Contextual Content to Context Ontology Process

II. RELATED WORK

Conceptual clustering, concepts are grouped according to the semantic distance between each other to make up hierarchies. But because of lack the domain context to instruct in the process of distance computation, the conceptual clustering process can't be efficiently controlled. Furthermore, by this method, only taxonomic relations of the concepts in the ontology can be generated [8].

Concept learning, a given taxonomy is incrementally updated as new concepts are acquired from real-world texts. Concept learning is a part of the process of ontology learning [9].

Text2Onto [10] creates an ontology from annotated texts. This system incorporates probabilistic ontology models (POMs). It shows a user different model ranked according to the certainty ranking and does linguistic preprocessing of the data. It also finds properties that distinguish a class from another.Text2Onto uses an annotated corpus for term generation.

Association rules, the association rules have been used to discover non-taxonomic relations between concepts. Association rules are most used on the data mining process to discover information stored on database. Ontology learning mostly uses unstructured texts but not the structure data in database. So, association rule is just an assistant method to help the ontology generation [11].

OntoLearn [12] uses an unstructured corpus and external knowledge of natural language definitions and synonyms to generate concepts.. But the ontology that is generated is a hierarchical classification and does not involve property assertions.

Wen Zhou [13] have proposed a semi-automatic technique that starts form small core ontology constructed by domain experts and learns the concepts and relations by use of the general ontology .In his paper ,WordNet and event based NLP technologies are used that automatically to construct the domain ontology.

H. Kong [14] gave the methodology for building the ontology automatically based on the frame ontology from the WordNet concepts and existing knowledge data. The ontology building method is divided into two parts. One part is to make the possibility for building the ontology automatically based on the frame ontology from the WordNet concepts that are the standard structured knowledge data.

Iqbal [15] proposed a semi automated algorithm to transform data to the ontology language, OWL. They described Ontology, as the vocabulary and core component of the Semantic Web, provides a re-usable representation of real-world things in a particular domain or application area.

In Sowa's Top Level Ontology [16], the ontology has a lattice structure where the top level concept is of universal and low level concepts are absurd type.

Wordnet [17] is the largest lexical database for English. It is divided into synsets each representing one lexical concept. Wordnet represents lexical entries into five categories. It is used by natural language processing based applications.

Maedche and Staab [18] distinguished different ontology learning approaches focus on the type of input used for learning, such as semi-structured text, structured text, unstructured text. In this sense, they proposed the following classification: ontology learning from text, from dictionary, from knowledgebase, from semi-structured schema and from relational schema

OntoEdit [19] provides an environment for the development of ontologies .It provides a user interface. The concept hierarchy can be edited or created .The decision of making direct instances of a concept depends upon the type of the concept.

OntoLT [20] allows a user to define mapping rules which provides a precondition language that annotates the corpus. Preconditions are implemented using XPATH expressions and consist of terms and functions. According to the preconditions that are satisfied, candidate classes and properties are generated. Again, OntoLT uses pre-defined rules to find these relationships.

III. Proposed Method for developing Ontology Using Information Extraction

This approach is to develop a method to develop context Ontology using information extraction approach for contextual content. The relationship between IE and ontologies can be considered in two non independent manners [1].

1. As IE can be used for extracting ontological information from documents, it is exploited by ontology learning and population methods for enriching ontologies.

2. And how ontologies can be exploited to interpret the document for IE purposes.

This paper focuses on how Information Extraction can be used for extracting ontological information from documents. Ontology is built linking concepts and relations extracted. After this we derive the context of a statement and add it to target Ontology. The final ontology is presented in the form of Context ontology. The proposed work will focus in particular on IE Contextual Assessment types that allow us to model rich Ontology adequately. The proposed method is shown in Fig.2.

CRCTOL [21] Concept Tuple based Ontology Learning (CRCTOL) its mines semantic knowledge in the form of ontology, in this paper a novel system, known as Concept Relation was introduced. By using a parsing technique and using statistical and lexico-syntactic methods, the knowledge extracted by their system is contains semantics compared with alternative systems..

J. Wang [22] used rule-based information extraction as a method to learn ontology instances. It automatically extracts the wanted factors of the instances, with the help of the definition in domain ontology.

Wu yuhuang [23] proposes a web based ontology learning model. This approach concerns realizing the ontology's automatic extraction from the Web page and exploring the pattern and the relations of the ontology semantics concept from the Web page data. It semi-automatically extracts the existing ontology through the analysis of Web page collection in the application domain.

Q. Yang [24] presents an Ontology Learning method which combines personalized recommendation with concept extraction and stable domain concept extraction method. This method uses machine learning for extraction of field concept. Recommendation study is used to domain concept extraction. It largely improves the accuracy of the concept extraction and the stability.

A. *Information Extraction*

In this phase, the document is analysed for the purpose of identifying these properties using ontology. If the axioms found in document compared with properties of any concept in ontology, the document is explained with that concept. In this way, documents are listed according to these properties. The main objective of this work is to present approaches of how these axioms can be extracted from documents; both for the purpose of Information Extraction and ontology building. To achieve this, there are two different approaches to information extraction; "Rule Based Approach" and "Dependency Based Approach.

1) Rule-Based Information Extraction: In rule based information extraction approach, grammatical rule or pattern recognition techniques are applied to extract out ingredients, culinary actions and their relationships. For this purpose, on the basis of analysis of recipe text, few grammatical rules have been designed which capture the ingredient, action and their relation from the text.

2) *Dependency Based Information Extraction*: In this approach of information extraction, syntax analysis technique called based parsing is applied. In this technique the syntactic analysis of text is based on dependencies between words within a sentence. The syntactic structure of sentence is determined by the relation between a word and its dependents.

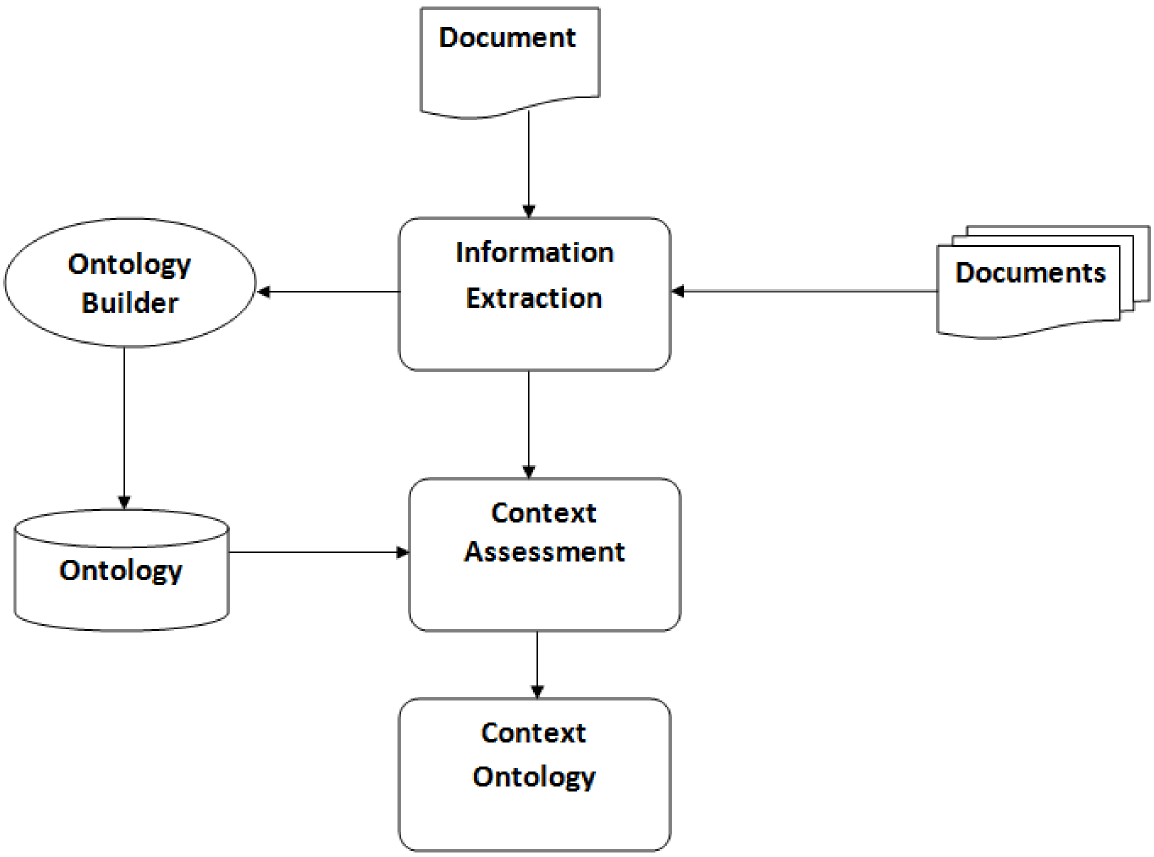


Fig.2: Proposed Framework for Developing Ontology using Information Extraction

B. Context Assessment Process

In this paper we focus on the scenario in which the target ontology is extended by introducing statements that have one part that already exists in ontology. The relation of can be either of taxonomic type or other named relations. Here we focus on the taxonomic relations, as they are less ambiguous than the named ones, of which relevance is complex to assess even by users. The Context assessment process starts with identifying a context C .The context C is matched to the target ontology.

It automatically selects and explores existing ontologies to discover relations between two given concepts. This process:-

1) Identifies existing ontologies that can provide information about how these two concepts interconnect.
2) Then Combines this information to infer their relation. To find existing ontologies in which statement appears, we use the subject and object of statement as input, which returns a list of relations that exist between the two objects, along with information about the source ontologies from where the relations have been identified.

In this process we find the linguistic terms that stand in place of other linguistics terms in the text. One of the example of context referent is "Bank". Here “Bank” refers to Bank of a river or a financial institute as shown in Fig .3. So there is a need to resolve this issue before processing.

C. Context Ontology Development

The ontology is constructed showing the relationships and the associated words in different context, with the help of the adding new concepts and relationships as in Fig.4.

IV. Conclusion

Since IE is an ontology-based activity and we suggest that future effort in IE should focus on formalizing and reinforcing the relation between the context extraction and the ontology model.

REFERENCES

[1] Claire N´edellec1, Adeline Nazarenko2, and Robert Bossy1 S. Staab and R. Studer (eds.), Handbook on Ontologies, International Handbooks on Information Systems, DOI 10.1007/978-3-540-92673-3

[2] Gruber, T. R. (1993). A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 5, Academic Press Ltd. 199–220.

[3] P.K. Bhowmick, D. Roy, S. Sarkar and A. Basu, A Framework For Manual Ontology Engineering For Management Of Learning Material Repository,2010.

[4] G. Aguado et al., "Ontogeneration: Reusing Domain and Linguistic Ontologies for Spanish Text Generation," Workshop on Applications of Ontologies and Problem Solving Methods (part of ECAI '98: 1996 European Conf. AI), European Coordinating Committee for Artificial Intelligence, pp. 1-10, 1998.

[5] Sowa, J. F. (2000). Knowledge Representation – Logical, Philosophical, and Computational Foundations, Pacific Grove, CA, USA: Brooks/Cole,2000.

[6] T. R. Gruber.(1995) Towards principles for the design of ontologies used for knowledge sharing. International Journal of Human-Computer Studies,43:907–928.

[7] Jaytrilok Choudhary and Devshri Roy ,An Approach to Build Ontology in Semi-Automated way, Journal Of Information And Communication Technologies, Volume 2, Issue 5, May 2012.

[8] D.Faure, T. Poibeau,(2000). "First experiments of using semantic knowledge learned by ASIUM for information extraction task using INTEX," In: S. Staab, A.Maedche, C. Nedellec, P. Wiemer-Hastings (eds.), Proceedings of the Workshop on Ontology Learning, 14th European Conference on Artificial Intelligence ECAI'00, Berlin, Germany.

[9] U. Hahn, S. Schulz.(2000). "Towards Very Large Terminological Knowledge Bases: A Case Study from Medicine," In Canadian Conference on Al 2000: 176-186.

[10] A. Maedche and S S. Staab.(2000). Semi-automatic engineering of ontologies from text. In 12th International Conference on Software Engineering and Knowledge Engineering.

[11] Alexander Maedche and Steffen Staab.(2001).Ontology Learning for the Semantic Web, T h e S e m a n t i c W e b, IEEE.

[24] Qing Yang,Kai-min Cai, Jun-li Sun, Yan Li.(2010). "Design Analysis and Implementation for Ontology Learning Model",2nd Inter-national Conference on Computer Engineering and Technology.

[12] R. Navigli and A. Gangemi P. Velardi.(2003). Ontology learning and its application.

[13] Wen Zhou, Zongtian Liu, Yan Zhao, libin Xu, Guang Chen,Qiang Wu, Mei-li Huang, Yu Qiang(2006). "A Semi-automatic Ontology Learning Based on WordNet and Event-based NaturalLanguage Processing", ICIA.

[14] H. Kong, M. Hwang and P. Kim.(2006). "Design of the automatic ontology building system about the specific domain knowledge", 8th International Conference on Advanced Com-munication Technology (ICACT).

[15] Ashraf Mohammed Iqbal, Abidalrahman Moh'd and Zahoor Khan.(2011). A Semi-AutomatedApproach To Transforming Database Schemas Into Ontology Language, IEEE.

[16] Sowa, J. F. (1995). Top-level Ontological Categories. International Journal of HumanComputer Studies,43, 669–685.

[17] Goerge, M. A. (1995). Wordnet: A Lexical Database for English, Communication of ACM,38, 39–41.

[18] JU Kietz, A Maedche, R Volz.(2000)."A Method for Semi-Automatic Ontology Acquisition from a Corporate Intranet," In: N. Aussenac-Gilles,B. Biebow, S. Szulman(eds) EKAW'00 Workshop on Ontologies andTexts. Juan-Les-Pins, France. CEUR Workshop Proceedings 2000,51:4.1-4.14. Amsterdam, The Netherlands Available: http://CEURWS.org/Vol-51.

[19] Sure, Y., Angele, J., Staab, S. (2002). OntoEdit: Guiding Ontology Development by Methodology and Inferencing. Proceedings of the 1st International Conference on Ontologies,Databases and Applications of Semantics for Large Scale Information Systems,1205–1222.

[20] Ontology Learning from Text: Methods, Evaluation and Applications, P. Buitelaar, P. Comiano, and B. Magnin, Eds. IOS Press Publication.

[21] Xing Jiang and Ah-Hwee Tan.(2005).Mining Ontological Knowledge from Domain-Specific Text Documents, Proceedings of the Fifth IEEE International Conference on Data Mining, IEEE.

[22] J. Wang, C. Wang, J. Liu and C. Wu.(2006). "Information Extraction forlearning of Ontology Instances", IEEE International Conference on Industrial Informatics.

[23] Wu yuhuang ,Li yuhuang,(2009)." Design and realization for ontology learning model based on web",IEEE.

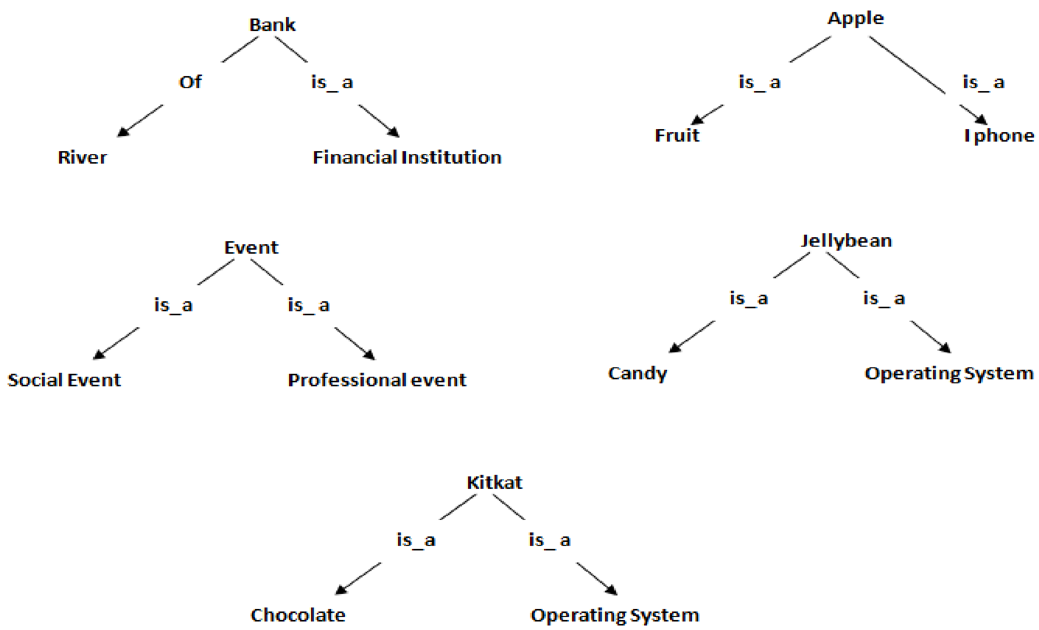


Fig.3: Proposed Framework for Developing Ontology using Information Extraction

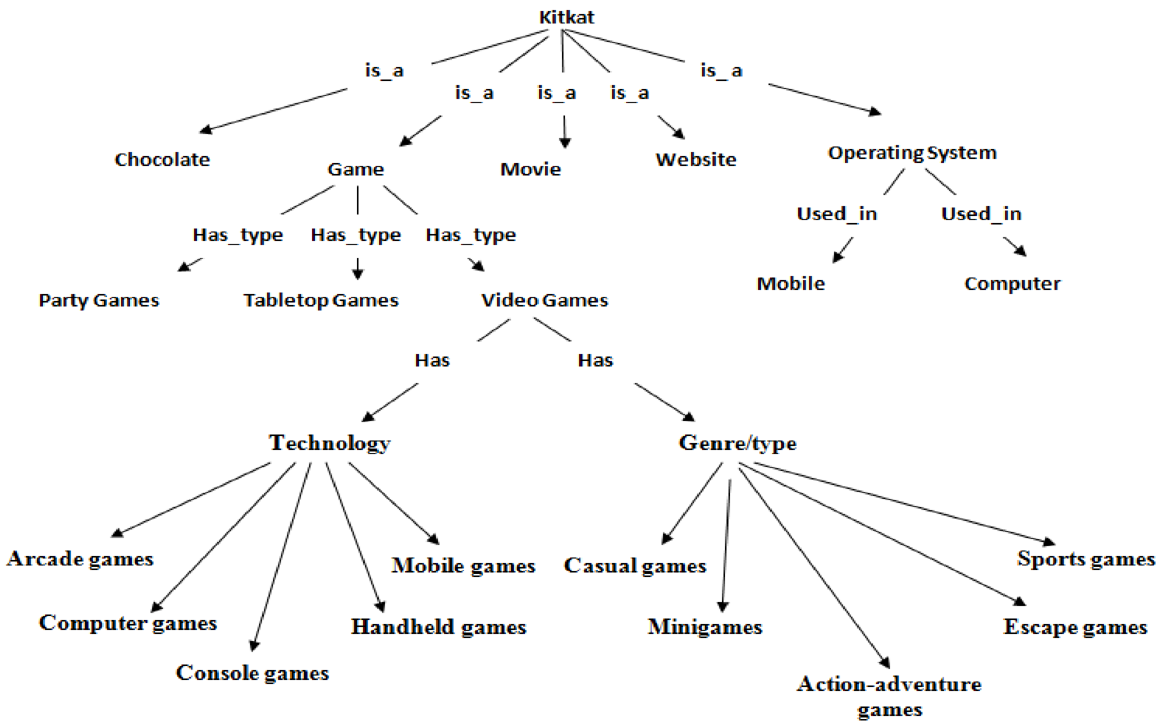


Fig.4: Context Ontological format of keyword

Challenges and Interesting Research Directions in Model Driven Architecture and Data Warehousing: A Survey

Amer Al-Badarneh, Jordan University of Science and Technology
Omran Al-Badarneh, Devoteam, Riyadh, Saudi Arabia

Abstract

Model driven architecture (MDA) is playing a major role in today's system development methodologies. In the last few years, many researchers tried to apply MDA to Data Warehouse Systems (DW). Their focus was on automatic creation of Multidimensional model (Start schema) from Conceptual Models. Furthermore, they addressed the conceptual modeling of QoS parameters such as Security in early stages of system development using MDA concepts. However, there is a room to improve further the DW development using MDA concepts. In this survey we identify critical knowledge gaps in MDA and DWs and make a chart for future research to motivate researchers to close this breach and improve DW solution's quality and performance, and also minimize drawbacks and limitations. We identified promising challenges and potential research areas that need more work on it. Using MDA to handle DW performance, multidimensionality and friendliness aspects, applying MDA to other stages of DW development life cycle such as Extracting, Transformation and Loading (ETL) Stage, developing On Line Analytical Processing(OLAP) end user Application, applying MDA to Spatial and Temporal DWs, developing a complete, self-contained DW framework that handles MDA-technical issues together with managerial issues using Capability Maturity Model Integration(CMMI) standard or International standard Organization (ISO) are parts of our findings.

Keywords: Data warehousing, Model driven Architecture (MDA), Platform Independent Model (PIM). Platform Specific Model (PSM), Common Warehouse Metamodel (CWM), XML Metadata Interchange (XMI)

---

1. INTRODUCTION

MDA (Model Driven Architecture) is a framework for software development proposed by the Object Management Group (OMG) and its models are important in the software development process. The software development process within MDA is driven by the activity of software systems modeling. Query View Transform (QVT) language is the OMG standards for performing the concept of transformation [OMG 2003b] [OMG 2008]. MDA can bring many benefits to all stakeholders of the organizations that adopt it in terms of production, cost, and platform independence. For management, MDA means more productivity has been gained by using MDA [MIKKO 2005]. MDA allows the time to complete new projects and the time to make changes on existing applications to shrink radically. Reducing the time will reduce the overall projects' cost.

MDA can reduce costs and time in many phases of system development. During requirements gathering, analyst places all information in a UML formal model, so analysts save time on requirements gathering and the information is automatically in the needed format. During design stage, designer can save time by turning analyst's UML models into more complete, precise and detailed models that are ready for code generation. MDA also allows programmers to save more time by re-using existing mappings to produce new applications, which will save the organization even more time. During testing stage, testers can produce testing scripts from the UML models without the need to write them. System supporter will benefit from the quality of MDA documentation because all changes are traced in UML model and everything is generated from these models, so updates are easy and fast [BEDIR et al. 2007].

MDA can increase Portability and platform neutrality because MDA focuses on modeling the system using platform independent models which can be easily transformed into new platforms by simply find or write mappings for the desired platform. MDA also allows for higher quality because most of MDA code is derived from a Platform Specific model (PSM) model, the possibility for human error is greatly minimized [MIKKO 2005].

Regarding Data Warehousing, [KIMBALL 2002] defined a Data Warehouse as "a copy of transaction data specifically structured for query and analysis". From this definition we can conclude that the DW is subject-oriented database that provides access to a consistent, historical version of organizational data using a set of tools to query, analyze, and present information in a way that helps in reaching a cost-effective business reengineering [KIMBALL 2002] [INMON 2005]. The ultimate goal for the DW is to use it to for business intelligence (BI) purposes. BI is the process of using the data in an enterprise to make the enterprise perform more intelligently. This is achieved by analyzing the available data, finding trends and patterns, and acting upon these. A typical BI application shows analysis results using “cross tab” tables or graphical constructs such as bar charts, pie charts, or maps [TORBEN 2009].

Because MDA has its distinct advantages over traditional Software development approaches, DW researchers try to apply MDA concepts on DW development layers. As stated by [JOSE-NORBERTO et al. 2005], DW has four layers: source layer, integration layer, customization layer and application layer. Based on this layered-architecture, MDA can be used to construct each layer. But to what degree MDA has been applied to these layers and what layers still need more MDA work; what types of DWs (Spatial, Traditional, Temporal) still have a space for more MDA painting; have NFR quality of service (QoS) requirements (e.g. security and performance requirements) been addressed using MDA concepts. Are there Software Process Improvement (SPI) initiatives, to create an MDA-related DW software engineering process that describes how to use MDA while developing the DW project and handle MDA technical issues together with management issues?

To give answers on these questions to better understand this multidisciplinary research field, we compiled research trends in MDA and DW. The compiled research is analyzed to discover the current and future research direction in these areas. The rest of the paper is organized as follows: Section 2 presents state of the art for DW and MDA; Section 3 represents our findings which show the current research direction in DW and MDA. Open problems are presented in Section 4, Future research trends are highlighted in Section 5. Finally, Section 6 summarizes the main conclusions.

2. STATE OF THE ART

This section contains the background knowledge about the state of the art for MDA and DW. Before applying MDA to DW, an overview of DW and MDA concept and architecture is needed to put our hands on the relevant areas of DW that can be developed using MDA. This overview will help readers, especially who are new to DW and MDA technology, to understand and recognize how MDA is used to build DW projects.

2.1 Data Warehouse (DW)

[INMON 2005] defines Data Warehouse term as “A Data Warehouse is a subject-oriented, integrated, time-variant, nonvolatile collection of data in support of management’s decisions”. From this definition, four key fundamentals for DW can be seen:

1) Subject orientation: the development of the DW is carried out in order to satisfy requirements of managers that will query the DW for specific business activities. The subject under investigation may be analyzing types of criminal cases at a given period base on location, analyzing product, etc. [ELZBIETA et al. 2004a].
2) Integration relates to the problem that data from different external and operational systems have to be joined. In this process, some problems have to be resolved: differences in data format, data codification, homonyms, synonyms, multiplicity of data occurrences, nulls presence, default values selection, etc. [PETER 2005].
3) Non-volatility implies data durability and stability: data can neither be modified nor removed [RACHID 2001].
4) Time-variation: [ELZBIETA et al. 2006] defines time variation as the ability to remember historic facts and perspectives. It is imperative to be able to know how something was classified or who owned something and how this changed over time. Moreover, it indicates that we may count on different values of the same object as it progresses over time.

2.1.1 DW Architecture

Figure 1 shows a 4-tier Data Warehouse architecture. This architecture aims to help technical staff to reach a Data Warehouse system that helps in decision making and utilizing large pools of data sources that the enterprise has, by converting these pools into valuable source of information. Data Warehouses are constructed in a sequential manner, where one phase of development depends entirely on the results attained in the previous phase. First, data is loaded into the DW. It is then used and analyzed by the Business analyst. Next, and after reviewing the feedback from the end user, the data is modified and/or other data is added. Then another portion of the Data Warehouse is loaded, and so forth. This iterative and feedback loop continues throughout the entire life of the Data Warehouse [INMON 2005]. A detail description about each DW tier will be provided in the following subsections

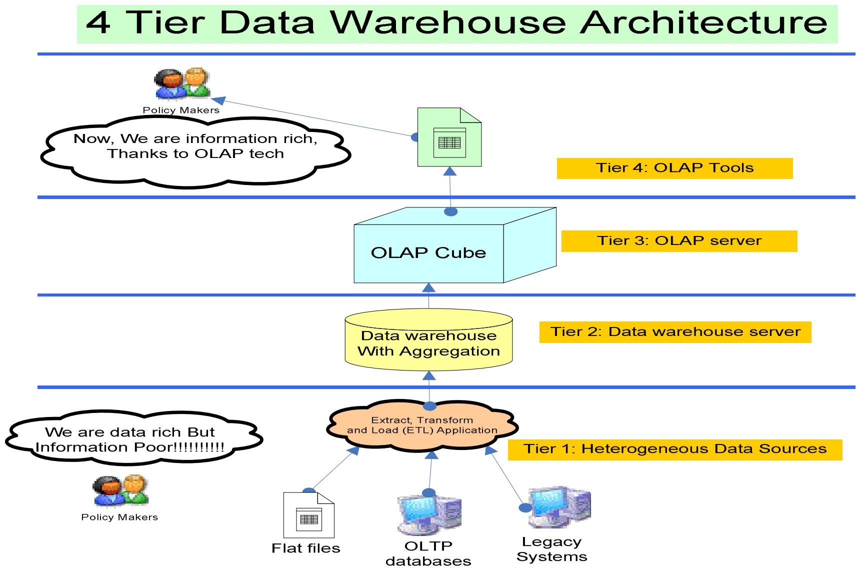


Figure 1. 4-Tier Data Warehouse Architecture.

2.1.1.1 Tier 1: Heterogeneouse Data Sources

This tier represents all different types of data sources of operational environment that will be moved to the DW server. The operations associated with moving and integrating these data sources are considered the most challenging and time consuming activities. In DW literature, these operations are called "Extracting, Transforming and loading (ETL)". Design begins with the considerations of placing data from unintegrated applications in the Data Warehouse server [KIMBALL 2002].

There are many considerations to be made concerning the placement of data into the Data Warehouse from the operational environment. First one is the integration of existing legacy systems. Second one is identifying types of loads and refresh approaches that will be used to keep the DW up to date. In other words, we have to take care of the efficiency of accessing existing systems data to avoid loading a file that has been loaded previously [LILIA et al. 2008]. Regarding Integration of existing legacy systems, most of legacy systems and data source in the operational environment are unintegrated. Lack of integration in existing systems is a fact. When the existing applications were developed, no thought was given to possible future integration. Each system had its own set of unique and private requirements [INMON 2005].

Pulling data from many applications or systems, integrating, and unifying it into a consistent, unified picture is a difficult task. This lack of integration is the nightmare for DW team. Many programming details must be taken into consideration just to pull the data properly from the operational environment [KIMBALL 2002]. A major problem is the efficiency of accessing existing systems data. How Extract-Transform-Load (ETL) tools or programs avoid scanning already scanned data? The existing data sources hold large amount of data, and attempting to rescan all of already scanned data when a Data Warehouse load needs to be done is inefficient and impractical [KIMBALL 2002].

Loading data on an ongoing basis as changes are made to the operational environment presents the largest challenge to the DW team. There are four solution used to avoid rescanning the data files at the point of refreshing the Data Warehouse. The first technique is to scan time stamped data in the operational environment. The second solution is to scan a delta file. The thirds solution to limiting the data to be scanned is to use the database log file. The fourth solution is to modify application code to directly deal with the Data Warehouse tables [INMON 2005].

2.1.1.2 Tier 2: Data Warehouse Server

The Data Warehouse server in general is a relational database engine which holds the MultiDimensional start schema, Figure 2 shows a conceptual star schema which is composed of fact table and set of surrounding dimension tables [IL-YEOL et al. 2008] [LYNN 2007]. From business point of view, this server represents the dimensional analysis view of the organization. When senior leadership and top management start the performance analysis of an organization, for example, to analyze the achievement of sales department, they have to look at the factors that influence the profit in the region.

These factors are likely to be sales team performance, products sold, customers and channels used for selling customers over time. As we can see, such situations can be thought of as problems with many attributes and dimensions. For sales managers, they want to improve the sales amount and profit of the region; so the dimensions of the problem are products, sales staff, distribution channels, regions and time. Any business analyst using data warehousing to sort out a given problem will work on a problem that has multi dimensions [MARK 2009].

2.1.1.3 Tier 3: OLAP Server

The previous tier as mentioned earlier is a traditional database server saving the data in star schema with a fact table surrounded by dimension tables. This type of data structure is not efficient for Multidimensional data analysis that provides quick answers with summarization or aggregation on dimension or executing roll up or drill down operations. The better data structure is an array based one which is supported by OLAP server [KRZYSZTOF et al. 2007]. To decrease the query time and to provide different viewpoints for the business users, these data are usually organized as data cubes. Each cell in a data cube corresponds to a unique set of values for the different dimensions and contains the measures [WEN et al. 2004].

Figure 3 presents an example of data cube. Data cube is structured by multiple dimensions such as products, customers and suppliers; and measures such as unit sold and average price. Dimensions may have one or more members (individual customers, product categories, sales territories) [WEN et al. 2004]. Dimensions are structured into one or more hierarchies. Hierarchies specify how data at the bottom level rolls up and might include levels (product, product group, product category), and attributes are used to describe characteristics and features of a dimension member, such as the color, size, or product code [MARK 2009].

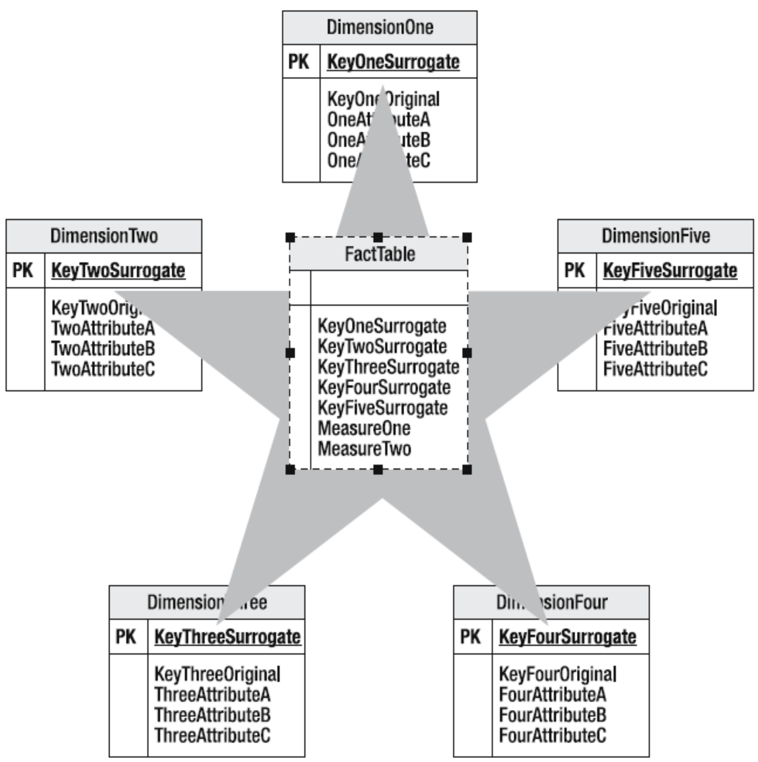


Figure 2. Basic concept of Star Schema [LYNN 2007].

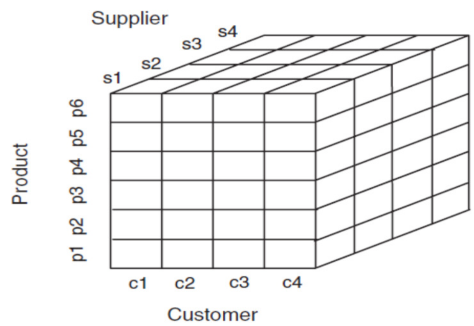


Figure 3. An example of data cubes [WEN et al. 2004].

2.1.1.4 Tier 4: OLAP Tools

OLAP Tools use and utilize the OLAP server components such as cubes, dimensions, measures, hierarchies, levels to provide managers with online analytical facilities such as rollup, drill down, slice, dice operations [KRZYSZTOF et al. 2007]. With OLAP support, the sales figures that the sales managers are trying to recognize are generated by various interactions between products, customers, and supplier over time. Sales managers need to think MultiDimensionally and OLAP tools present data to users in a way that mirrors this MultiDimensional way of thinking [MARK 2009]. Figure 4 depicts an example screenshot of the OLAP tool sold by the Danish company TARGIT. Such BI solutions are typically very easy to use, and can be used by non-technical business people like business analysts or managers [TORBEN 2009].

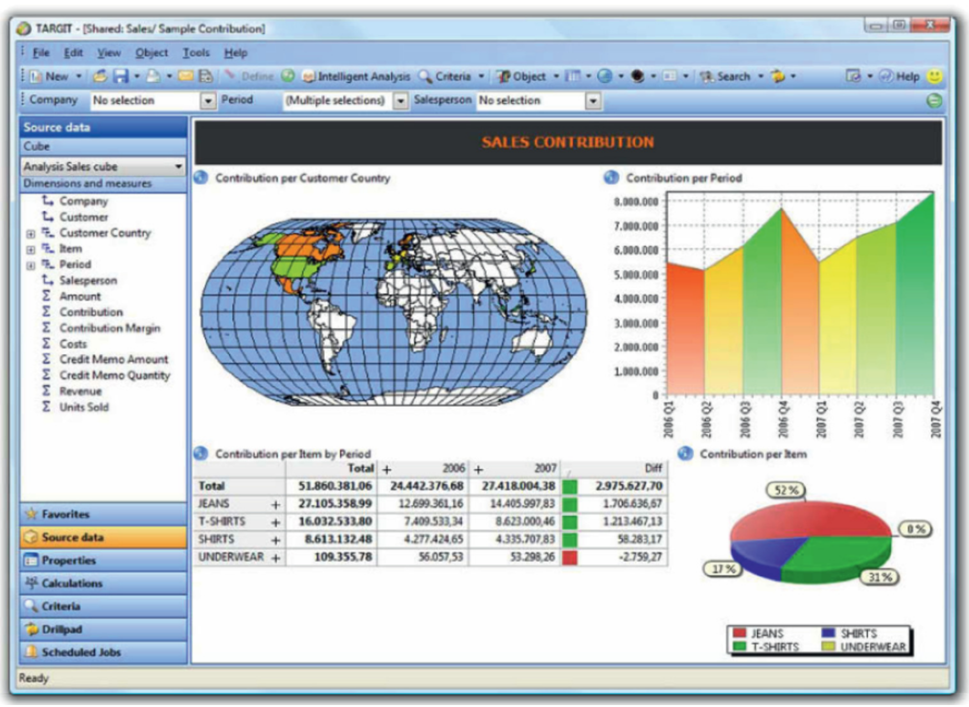


Figure 4. Example of OLAP Application [TORBEN 2009].

2.2 Model Driven Architecture (MDA)

MDA is a software development lifecycle. It models development artifacts [OMG 2003b]. The level of abstraction in software engineering is raised to develop complex applications in simpler ways. The system functionality is separated from the implementation details. So, MDA is language, vendor and middleware neutral [SERGIO et al. 2006], [JOSE-NORBERTO et al. 2005].

MDA focuses on the modeling task and transformation between models. Figure 5 presents the layout of MDA framework. Firstly, we build a Computational Independence Model (CIM) that describes the system within its environment and its business domain. The model shows what the system is expected to do but without showing details about how it is constructed. Therefore, requirements of the system are modeled by a Computation Independent Model. Then, these requirements are traceable to the PIM (Platform

Independent Model) and PSM (Platform Specific Model) that realize them [OMG 2003b]. CIM can be modeled by using UML as well using "use cases diagrams" or other languages more specific to requirement analysis, such as i* [YU 1997].

After building the CIM model, we build the PIM model. Then, we automatically create the PSM out of the PIM using Model-2-Model (M2M) transformation language. PSM can be in any proprietary platform we want (e.g. RDBM, CORBA, J2EE, .NET, XMI/XML). When PSM is finished, the code of the system can be generated from the PSM model using Model-2-Text (M2T) language. The generated code is illustrated in Figure 6 [OMG 2003b]. PIM and PSM can be constructed using any specification language. UML, a standard modeling language is typically used and easily extended to define specialized languages for certain domains [OMG 2009].

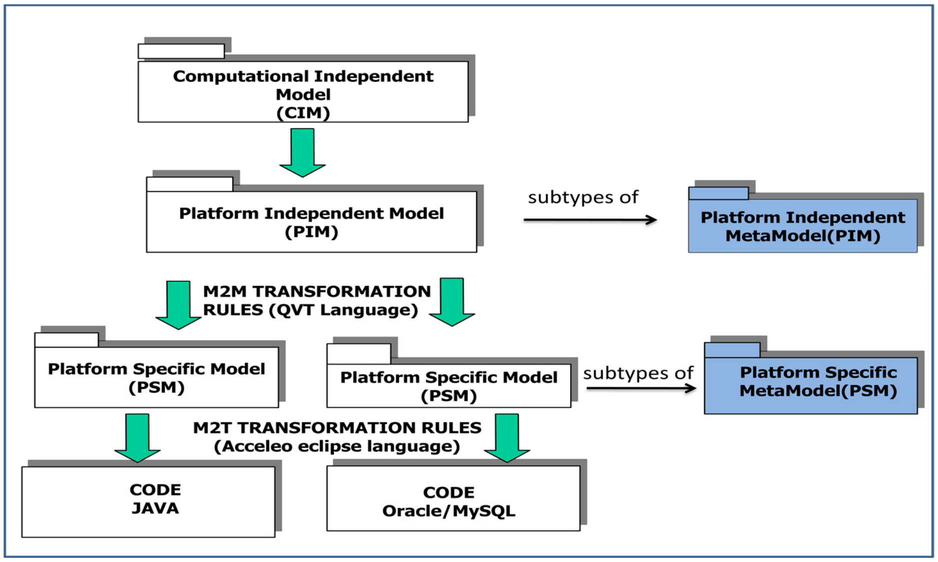


Figure 5. Model-Driven Architecture framework [JOSE-NORBERTO et al. 2005]

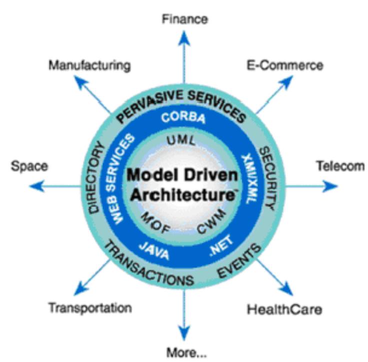


Figure 6. OMG Model-Driven Architecture Model [OMG 2003b].

There are four principles that show the OMG's MDA approach [OMG 2003b]:

1) Models expressed in a well-defined notation are a cornerstone to system understanding for enterprise-scale solutions.
2) Building systems can be organized around a set of models by imposing a series of transformations between models, organized into an architectural framework of layers and transformations.
3) A formal foundation for describing models in a set of metamodels that facilitates meaningful integration and transformation among models, and is the basis for automation through tools.
4) Acceptance and broad adoption of this model-based approach requires industry standards to provide openness to consumers, and faster competition among vendors.

2.2.1 The Key Standards of MDA

The key standards of MDA are:

1) UML is a graphical language for visualizing, specifying, constructing and documenting the artifacts for software systems and can be used for designing models in PIM [OMG 2009].
2) Meta Object Facility (MOF) is an integration framework for defining, manipulating and integrating metadata and data in a platform independent manner. It is the standard language for expressing metamodels. A metamodel uses MOF to formally define the abstract syntax of a set of modeling constructs [JOHN et al. 2003], [OMG 2008].
3) XML Metadata Interchange (XMI) is an integration framework for defining, interchanging, manipulating and integrating XML data and objects. XMI can also be used to automatically produce XML Document Type Definitions (DTD) and XML schemas from UML and MOF models [OMG 2007].
4) Common Warehouse Metamodel (CWM) is a set of standard interfaces that enable easy exchange of business intelligence metadata between Data Warehouse tools, warehouse platforms and warehouse metadata repositories in distributed heterogeneous environments [OMG 2003a]. As explained by [LEOPOLDO et al. 2005] CWM is a language created specifically to model database applications. All the metamodels of CWM (Relational, OLAP, and XML) can be used as source and target in the transformation process [OMG 2003a].

3. CURRENT RESEARCH DIRECTION IN DW AND MDA

The survey browsed and reviewed the related research articles. These articles are categorized in two major classes; first one, which is non-MDA DW articles, focuses on DW requirement elicitation frameworks. The other one is an MDA-related DW articles, which apply MDA concepts to DW development phases.

3.1 DW Requirement Elicitation Frameworks

This section presents set of papers that addressed DW requirements elicitation process for Functional Requirements and Non-Functional Requirements. We see this is important because requirement elicitation is a fundamental step before starting modeling the DW requirements, whether we will use MDA or other approaches to model the requirements. Moreover, this will help us understand how MDA can conceptually model both Functional and Non-Functional Requirements at early stages of the DW project.

3.1.1 Functional Requirement Elicitation Frameworks

[WINTER et al. 2003] proposed a broad methodology that supports the entire process of developing Functional Requirements of Data Warehouse projects, matching Functional Requirements with actual data sources, evaluating the collected Functional Requirements, specifying priorities for unsatisfied Functional Requirements, and formally identifying the results as a basis for incoming phases of the Data Warehouse development project.

Because Requirement analysis may involve significant problems if it conducted in faulty or incomplete way, requirements analysis should attract particular attention and should be comprehensively supported by effective methods. Hence it is fair to assume that the special nature of DW systems justify a specific methodology for requirement analysis [WINTER et al. 2003]. Figure 7 depicts the activities involved in the methodology proposed by WINTER et al. 2003].

[NAVEEN et al. 2004] introduced a new requirement elicitation process for DWs by identifying the goals of the decision makers and the required information that supports these goals. This elicitation process can identify DW information contents that support set of decisions made by business owners. [NAVEEN et al. 2004] proposed an Informational Scenario as the means to elicit information for a decision. An informational scenario is written for each decision and is a sequence of pairs of the form <Query, Response >. A query requests for information required to take a decision and the response is the information itself. The set of responses for all decisions make out DW contents [NAVEEN et al. 2004].

[PAOLO et al. 2005] presented a goal-oriented framework to model requirements for DWs, thus obtaining a conceptual MD model from them by using a set of guidelines. In this paper two different perspectives are integrated for requirement analysis: organizational modeling, centered on stakeholders, and decisional modeling, focused on decision makers. This approach can be employed within both a demand-driven and a mixed supply/demand-driven design framework.

This goal-oriented framework can help the DW designer to reduce the risk of project failure by ensuring that early requirements are properly taken into account and specified, which ensures a good design and that the resulting DW schema is tightly linked to the operational database which makes the design of ETL simpler [PAOLO et al. 2005].

Discussion:

The previous mentioned works ([WINTER et al. 2003], [NAVEEN et al. 2004], [PAOLO et al. 2005]) have considered requirement analysis as a crucial task in early stages of the DW development process. However, these approaches have the following disadvantages: (i) only consider Functional Requirements, but Non-Functional Requirements have not been articulated. (ii) These approaches are not part of a complete methodology. They need some kind of integration and adaptation to be used in an existing corporate methodology. (iii) These approaches do not describe how to model these requirements; just how to elicit them.

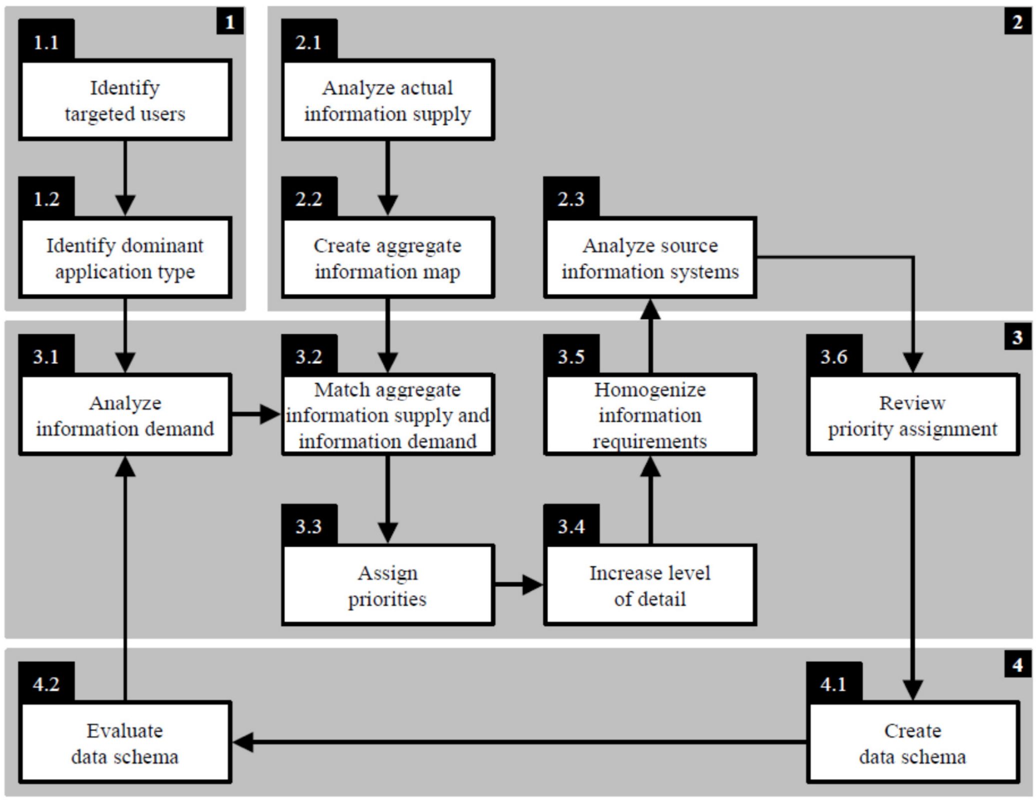


Figure 7. Activity model for the proposed methodology [WINTER et al. 2003].

3.1.2 Non-Functional Requirement Elicitation Frameworks

[PAIM et al. 2002] addressed the enhancement of Data Warehouse design by extending the Non-Functional Requirement (NFR) Framework proposed by [CHUNG et al. 2000]. Catalogues of major DW NFR types and related operational methods has been defined. [PAIM et al. 2002] highlighted the importance of set of quality factors such as integrity, accessibility, performance, and other domain-specific Non-Functional Requirements (NFRs) that governs the success of the DW project. Figure 8 depicts a broad catalogue of most important Data Warehouse NFRs created by [PAIM et al. 2002]. Figure 9 depicts several techniques proposed by [PAIM et al. 2002] to tackle the DW performance issue.

Discussion:

This effort proposed by [PAIM et al. 2002] is promising by addressing most of NFR for DW. However, it has these shortcomings: (i) the specification of these requirements is considered in an isolated way, without taking Functional Requirements into account. However, in order to obtain a conceptual multidimensional (MD) model that drives the development of a DW, which satisfies functional needs as well as QoS (Non-Functional) needs, both types of requirements should be gathered and modeled together, since they are related [EDUARDO et al. 2006a]. (ii) This effort is not part of complete DW methodology; it needs to be integrated and linked to existing DW Methodology. (iii) This effort focuses only on techniques for gathering NFR, but not on how to model these requirements.

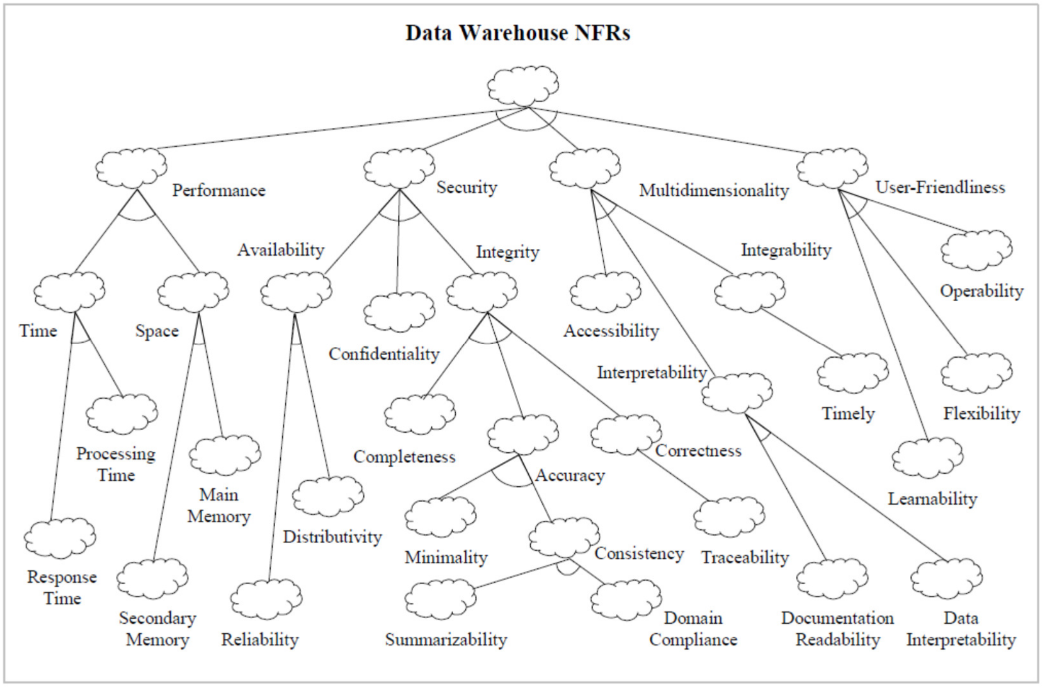


Figure 8. Hierarchical tree representing the Data Warehouse NFR Type Catalogue [PAIM et al. 2002].

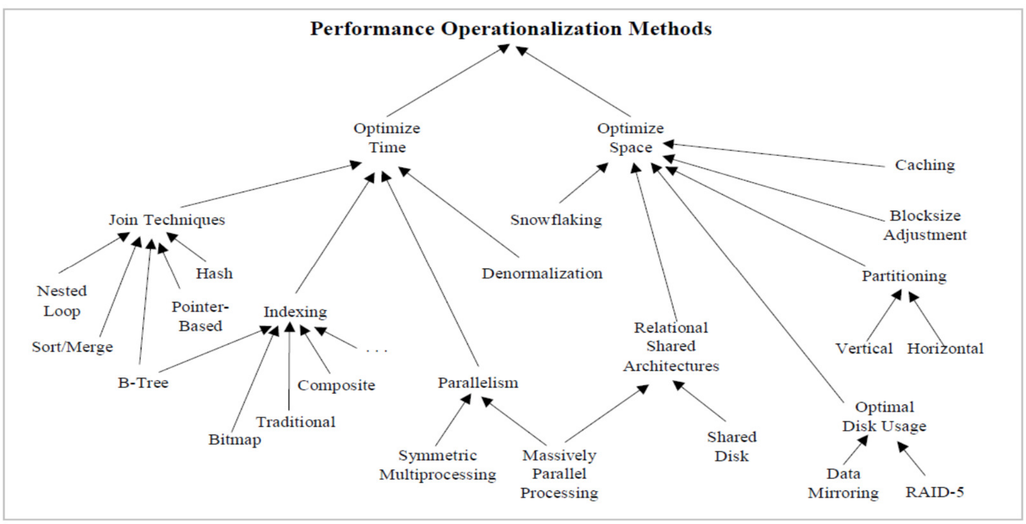


Figure 9. Performance Operationalization Catalogue [PAIM et al. 2002].

3.2 MDA-related DW Works

This section summarizes most of MDA-related works that has been conducted on DWs. These works are classified into the following groups

1) Creating CWM-based DW Modeling Tools.
2) Specifying Metamodel transformations for DW design.
3) Extending UML for MultiDimensional Modeling.
4) MDA-related DW framework.
5) Securing DW using MDA.
6) Designing Spatial Data Warehouse using MDA Techniques.
7) New Business-level Security UML profile.
8) Conceptual OLAP Platform-independent Queries.
9) MDA Framework for Designing Spatial DWs.

3.2.1 Creating CWM-based DW Modeling Tools

[KUMPON et al. 2003] presented a tool named ER2CWM that can be used to design relational databases with physical Entity Relational (ER) model and create database schemas by using Common Warehouse Metamodel (CWM). ER2CWM tool supports the creation of ER diagrams, transformation into CWM format, and creation of database schemas for relational database management systems. It can also transform database schemas back into CWM and ER diagrams respectively. With ER2CWM, database designers are assisted at database design, creation, and maintenance via a standard CWM format that can also be moved for use in other environments [KUMPON et al. 2003].

CWM is a standard XML-based metamodel for describing Data Warehouse models, allowing these models to be exchanged between heterogeneous environments in a convenient way [OMG 2003a, OMG 2009]. ER diagrams are generally used to express designs of relational databases. There are tools, such as OracleDesigner, PowerDesigner, and ERwin, that can help database designers to design a database with ER diagrams and create database schemas. All of these tools usually support the reverse database engineering to create ER diagrams from existing database schemas also [KUMPON et al. 2003].

All schemas in these tools are done via vendor-based schema representations that are specific to individual design tools. This means that each tool has its own metadata format that represents ER models and is used to create database schemas. Now exchanging models between different environments is not convenient for the designers to export a database schema designed and created by one tool to other working environments since a mapping process between the metadata of the source and target environment will be required for each pair of the exchanging environments [KUMPON et al. 2003].

OMG provides a solution to the vendor-based model exchange problem by standardizing the Common Warehouse Metamodel (CWM) for easy interchange of metadata between data warehousing tools and metadata repositories in heterogeneous environments [KUMPON et al. 2003] [OMG 2003a]. CWM is a specification for modeling metadata for relational, non-relational, MultiDimensional systems [JOHN et al. 2003]. As illustrated in Figure 10, CWM-based metadata are exchanged in the XML Metadata Interchange (XMI) documents [OMG 2009]. With CWM, metadata can be exchanged between data stores, warehouse builder tools, OLAP and end-user tools, and metadata repository and tools [KUMPON et al. 2003].

To incorporate CWM for simple exchange between database design tools, their specific metadata format has to be transformed to CWM layout using a tool called the Meta Integration Model Bridge (MIMB) [MIT 2003],[KUMPON et al. 2003]. This is a tool that can convert specific data of one design tool to another, including CWM. Figure 11 shows the exchange of the metadata between PowerDesigner and ERwin via CWM that is generated by MIMB. Since MIMB is a metadata converter, not a design tool, it is not convenient if there is a need to change the database schema that is in CWM format. Maintaining databases in these scenarios involve several steps and tools [KUMPON et al. 2003].

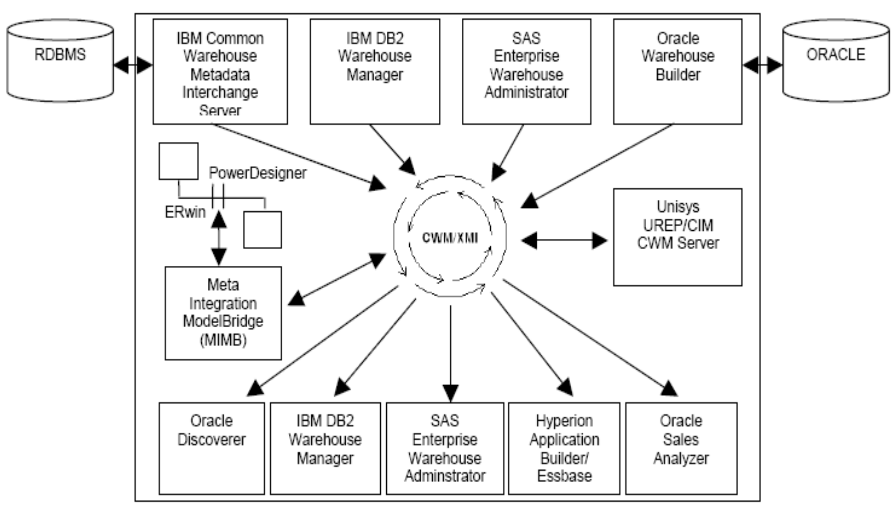


Figure 10. Interoperability via CWM [KUMPON et al. 2003].

To solve MIMB problems, [KUMPON et al. 2003] designed and developed the ER2CWM tool that can be used to create database schemas for particular database management systems by using CWM as its metadata format. The tool supports the design of the physical data model of the databases, generates CWM

Relational metadata, and creates database schemas. With this single tool, databases designers are provided with the power of a database design tool and a CWM converter by which maintenance of schemas is convenient and compliance with industry standard metadata format [KUMPON et al. 2003].

Discussion:

This [KUMPON et al. 2003] effort represents a solution for tool-based metamodel exchange problem by using a CWM-based tool which eases the process of metamodel transfer between different environments. With this good effort, there is a space for enhancing or building a tool to support all CWM specifications. This work focused only on the part of CWM for relational database schemas called CWM Relational while CWM is a specification for modeling metadata for Relational, Non-Relational, MultiDimensional systems.

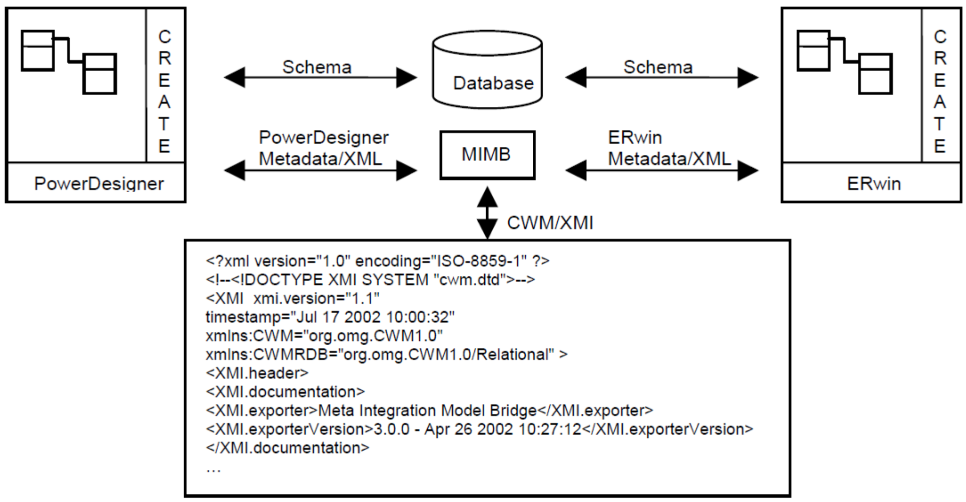


Figure 11. Use of database design tools and MIMB [KUMPON et al. 2003].

3.2.2 Specifying Metamodel Transformations for Data Warehouse Design

[LEOPOLDO et al. 2005] proposed an automatic approach for generating MultiDimensional structure (OLAP) from relational databases, based on MDA. The main contribution of [LEOPOLDO et al. 2005] work was a set of transformation rules used for developing tools that support the application of reusable transformations. MDA supports the development of software systems through the transformation of models. MDA requires that model transformations be defined precisely in terms of the relationship between a source metamodel and a target metamodel [ANNEKE et al. 2003]. [LEOPOLDO et al. 2005] succeeded in deriving OLAP schema (target metamodel) from the Relational metamodel (source metamodel). Figure 12 shows the OLAP target metamodel and Figure 13 shows the Relational metamodel.

Discussion:

[LEOPOLDO et al. 2005] highlighted a fundamental idea in MDA, which is the need for mapping between PMI source metamodels and PSM target metamodels. [LEOPOLDO et al. 2005] work can be enhanced by increasing the set of transformations to start from UML Conceptual model then generating Relational model to be converted to OLAP model. Moreover incorporating theses transformations into a methodology framework will add more value to targeted methodology.

3.2.3 Extending UML for MultiDimensional Modeling

[SERGIO et al. 2006] proposed a new UML profile by extending Unified Modeling Language for MultiDimensional modeling in Data Warehouses. The Developed UML profile was defined by a set of stereotypes, constraints and tagged values to elegantly specify main MD properties at the conceptual level. Object Constraint Language (OCL) has been used to specify the constraints attached to the defined stereotypes, thereby avoiding an arbitrary use of these stereotypes [SERGIO et al. 2006].

UML has been used as the modeling language for two main reasons: (i) UML is a well-known standard modeling language known by most database designers, thereby designers can avoid learning a new notation, and (ii) UML can be easily extended so that it can be tailored for any specific domain such as the MultiDimensional modeling for Data Warehouses [SERGIO et al. 2006]. Moreover, [SERGIO et al.

2006]'s proposal was an MDA compliant and they used the Query-View-Transformation language for an automatic generation of the implementation in a target platform.

Discussion:

This thesis considers the work developed by [SERGIO et al. 2006] is a valuable contribution to the MDA literature because it had given a clear example on how to extend UML2 and identified by example the main required elements to build an MDA-based project. These elements are: (i) UML profile that describes the all the concepts (stereotypes) that we will use in modeling the new line of business, (ii) PIM and PSM metamodels that are created based on the regarded profile and (iii) QVT (Query, View, Transformation) language to transform PIM to PSM.

For future work, [SERGIO et al. 2006] proposed extending the current UML profile to contain new stereotypes regarding object-oriented and object-relational databases for an automatic generation of the database schema into these kinds of databases. They also proposed extending the current profile for considering the conceptual modeling of secure Data Warehouses as well as considering the specification of dynamic aspects of the MultiDimensional modeling such as the modeling of end user requirements for the current profile version [SERGIO et al. 2006]

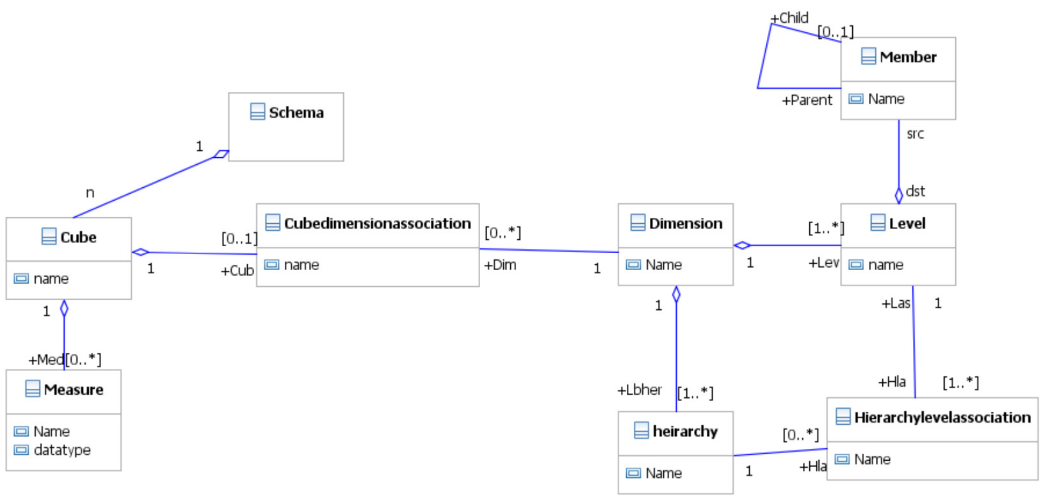


Figure 12. OLAP metamodel [LEOPOLDO et al. 2005].

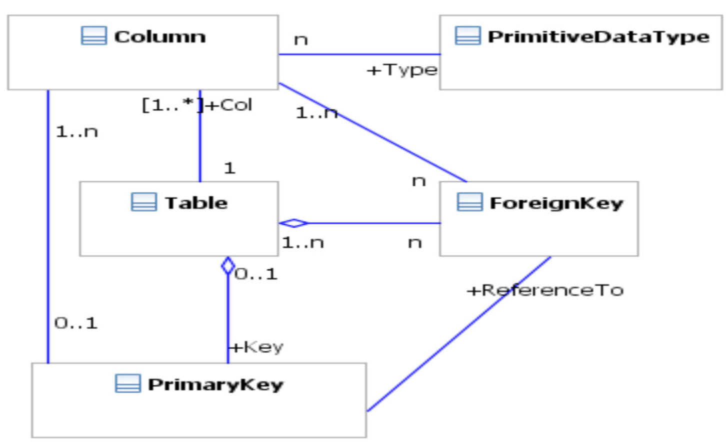


Figure 13. Relational metamodel [LEOPOLDO et al. 2005].

3.2.4 MDA-related DW framework

[JOSE-NORBERTO et al. 2005] presented a big picture for an MDA-related DW framework that aligns MDA standards with development of Data Warehouses. The architecture of a DW is usually presented as a multi-layer system in which data from one layer is derived from data of the previous layer as presented in Figure 14 [JOSE-NORBERTO et al. 2005].

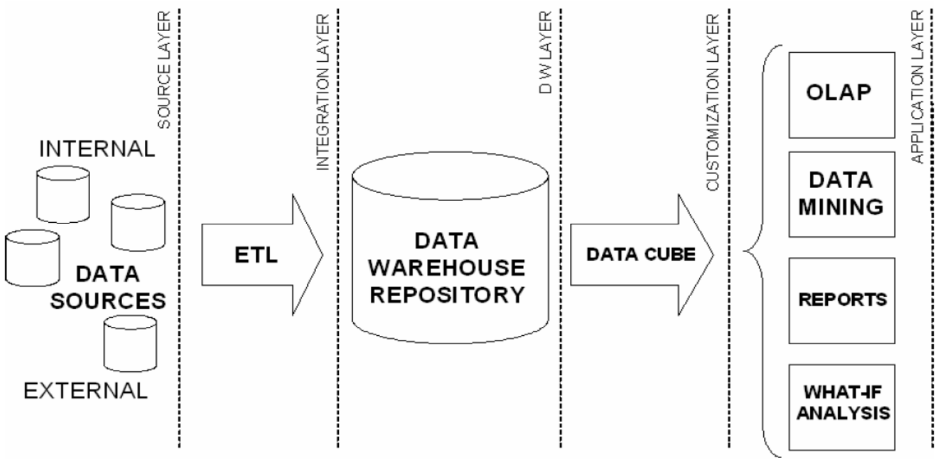


Figure 14. Multi-layer Data Warehouse [JOSE-NORBERTO et al. 2005].

As stated by [JOSE-NORBERTO et al. 2005], different approaches had been presented for designing such various parts of DWs, but the whole design of a DW is not dealt with in an integrated way. To sort out every design drawbacks, partial solutions to certain issues are such as Extraction-Transformation-Load (ETL) processes or MultiDimensional (MD) modeling. Problems due to partial solutions derived from interoperability and integration between layers may still arise [JOSE-NORBERTO et al. 2005].

On the other hand, the Model-Driven Architecture (MDA) is a standard framework for software development that tackles the complete life cycle of designing and developing applications by using UML models in application development. Hence [JOSE-NORBERTO et al. 2005] developed MDA Data Warehouse framework and described how to align the whole DW development process to MDA. Figure 15 depicts a big picture for this MDA framework.

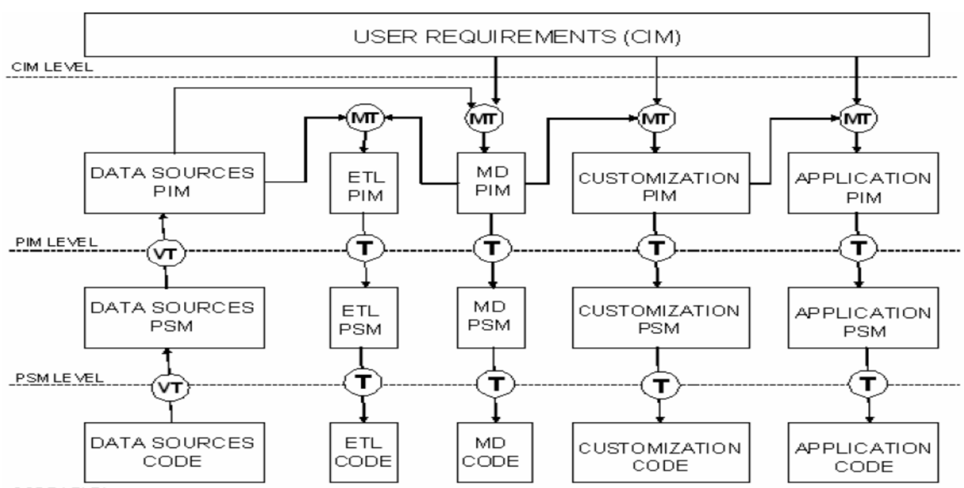


Figure 15. MDA oriented DW development framework [JOSE-NORBERTO et al. 2005].

They addressed the design of the whole DW system by aligning every layer of the DW with the different MDA viewpoints. Then, for each layer they had three different viewpoints (i.e. CIM, PIM, and PSM). The whole system was constructed by means of transformations applied to PIM in order to automatically obtain PSMs. From PSM, it was possible to obtain code in a straightforward way [JOSE-NORBERTO et al. 2005].

The development of the DW was reduced to create the PIM of each layer and its corresponding transformations. They have defined the MD PIM, the MD PSM and the necessary transformations. Their

PIM have been modeled by using their MD modeling profile [SERGIO et al. 2002b], [SERGIO et al. 2002a] as a PIM and the CWM Relational package [OMG 2003a] as a PSM, while the transformations were formally and clearly executed by using QVT [OMG 2005A]. Figure 16 depicts the MD PIM metamodel and Figure 17 depicts Part of the Relational CWM metamodel that had been used in model transformation process [JOSE-NORBERTO et al. 2005].

To give an example how to use MDA standard in building DW MD model, [JOSE-NORBERTO et al. 2005] developed a case study that used Oracle database as the target implementation engine. Figure 18 gives an abstract view that shows how conceptual MD PIM model is converted to MD PSM model and how SQL code is generated from the MD PSM model [JOSE-NORBERTO et al. 2005].

Discussion:

Even though [JOSE-NORBERTO et al. 2005] applied MDA just to MD layer, the work provided by [JOSE-NORBERTO et al. 2005] can be considered as a valuable contribution to the research field because it identifies the areas in DW development that can use MDA as its development method. How to use MDA for other layers of DW such as integration layer, Application layer and customization layer is an area that need more MDA work. Also applying MDA on different types of data warehousing project such as Spatial DW and Temporal DW is another open area for MDA improvement.

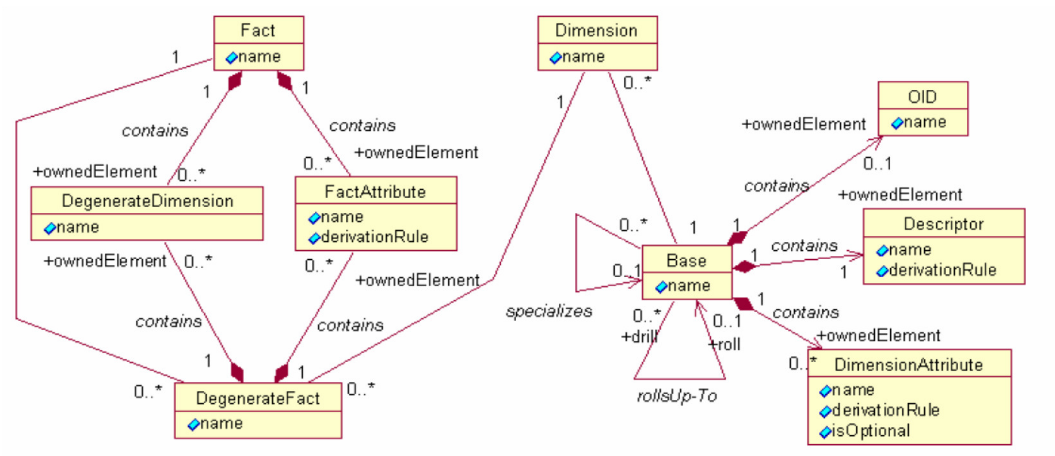


Figure 16. Metamodel used to design PIMs [JOSE-NORBERTO et al. 2005].

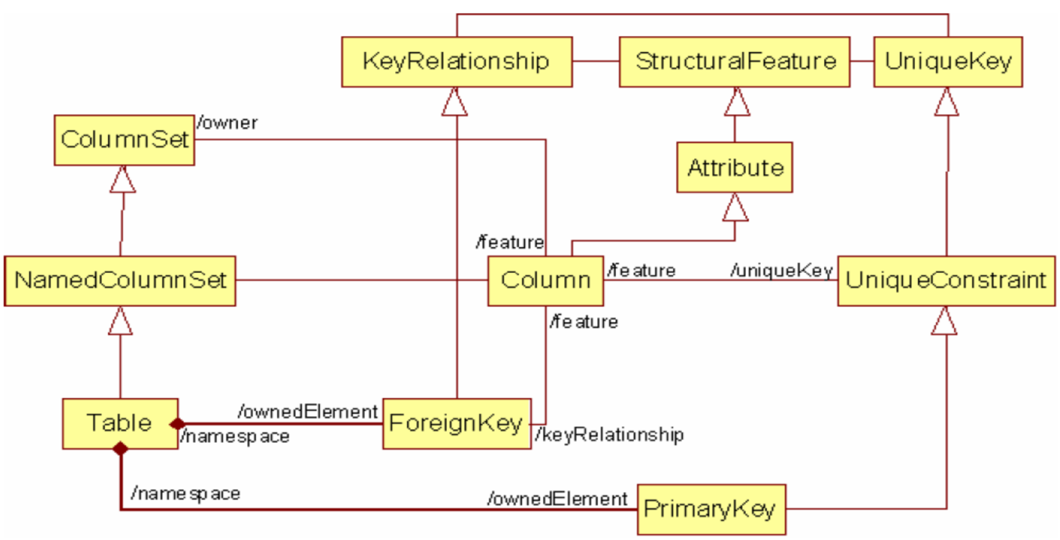


Figure 17. Part of the Relational CWM metamodel [JOSE-NORBERTO et al. 2005].

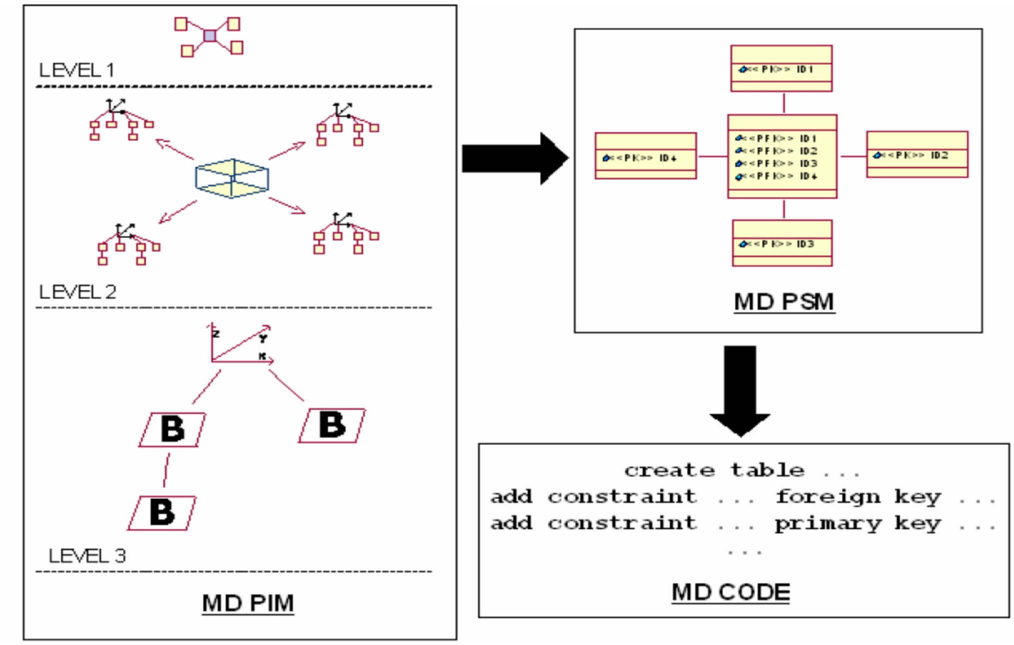


Figure 18. Obtaining MD PSM and code from a MD PIM [JOSE-NORBERTO et al. 2005]

3.3 Securing DW Using MDA

[RODOLFO et al. 2006] developed a new security profile named SECDW (Secure Data Warehouses) using the UML 2.0 extensibility mechanisms [OMG 2005b] [BRAN 2007]. This profile focused on solving confidentiality problems in the conceptual modeling of Data Warehouses. Figure 19 depicts a high level view of SECDW profile. In addition, [RODOLFO et al. 2006] defined an OCL [OMG 2003c] extension that allows specifying the security constraints of the elements in conceptual modeling of Data Warehouses and then applied this profile to an example.

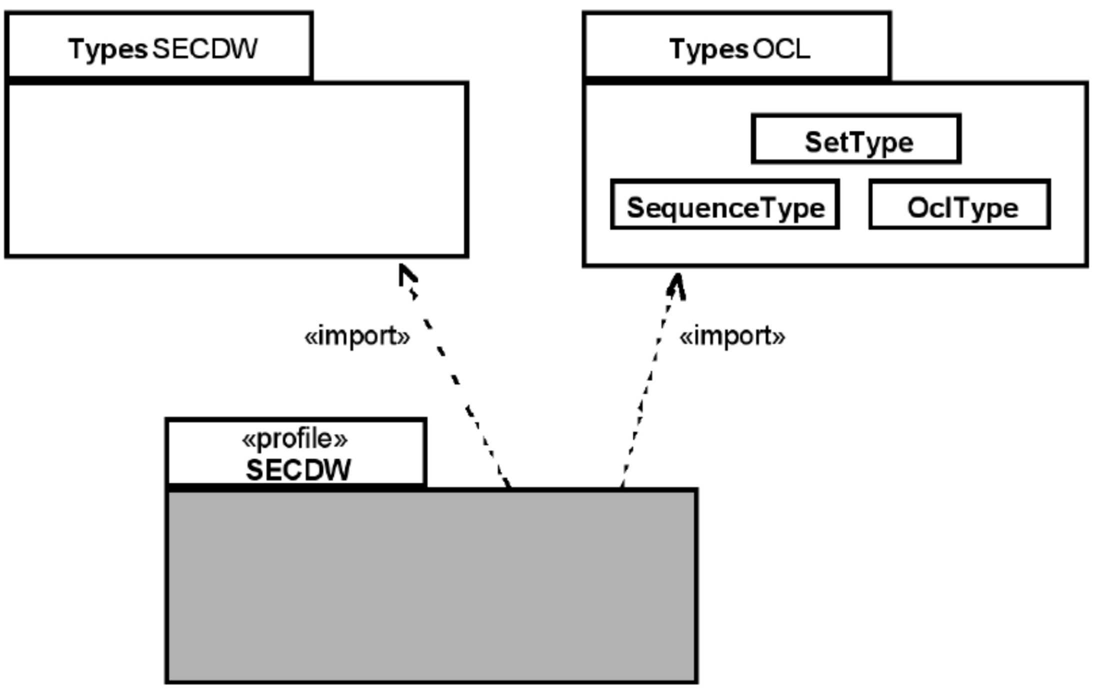


Figure 19. High level view of SECDW profile [RODOLFO et al. 2006].

[RODOLFO et al. 2006] explained the motivation and the need for this security profile. Security is a serious requirement that should be handled in proper way. Access control and multi-level security for DW was addressed after building the DW and security aspects were not considered during all system development life cycle (SDLC) stages as well as there were no approaches that introduced security into MD conceptual design [RODOLFO et al. 2006]. [RODOLFO et al. 2006] developed the security UML profile with final objective was to be able to design an MD conceptual model. They classified information in order to define which security properties the user had to possess in order to be entitled to gain access to information. For each element of the model (fact class, dimension class, fact attribute, etc.), its security information had been defined. They specified a sequence of security levels (multi-level security), a set of user compartment and a set of user roles. They specified security constraints considering these security attributes [RODOLFO et al. 2006].

SECDW profile will not only inherit all properties from the UML metamodel but it also incorporated new data types, stereotypes, tagged values and constraints. New data types are needed to be used in the tagged value definitions of the new stereotypes. Table 1 provides the new data type definitions that have been used in the profile and in

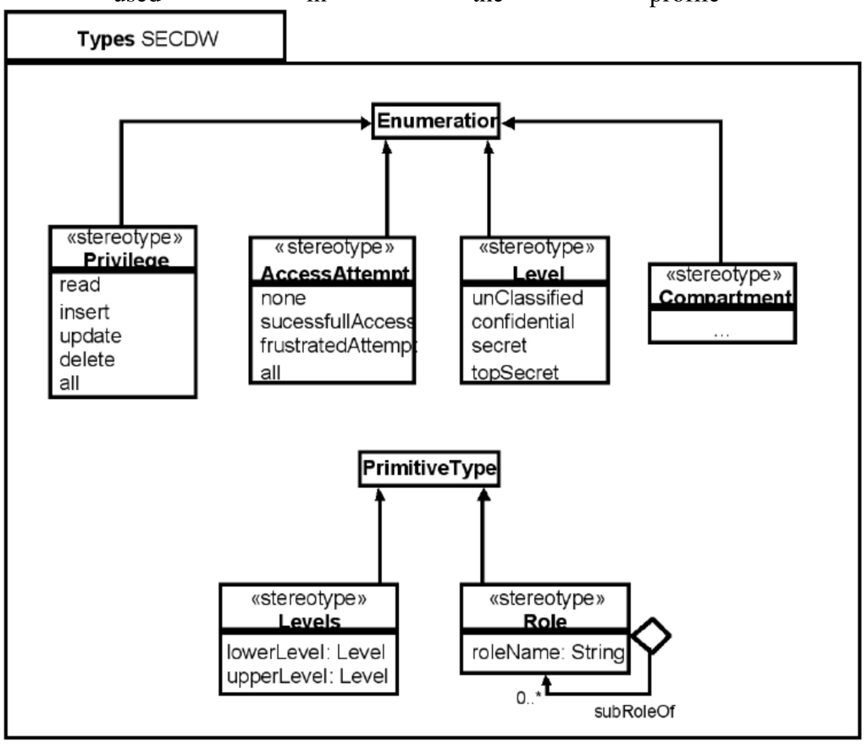


Figure **20**, the values associated to each one of the necessary data types can be seen [RODOLFO et al. 2006].

Table 1. New Data types [RODOLFO et al. 2006]

Name	Base class	Description
Level	Enumeration	The type Level will be an ordered enumeration composed of all security levels that have been considered.
Levels	Primitive	The type Levels will be an interval of levels composed of a lower level and an upper level.
Role	Primitive	The type Role will represent the hierarchy of user roles that can be defined for the organization.
Compartment	Enumeration	The type Compartment is the enumeration composed of all user compartments that have been considered for the organization.
Privilege	Enumeration	The type Privilege will be an ordered enumeration composed of all different privileges that have been considered.
AccessAttempt	Enumeration	The type Attempt will be an ordered enumeration composed of all different access attempts that have been considered.

All the information considered in these new data types has to be defined for each specific secure, conceptual database model, depending on its confidentiality properties, and on the number of users and complexity of the organization in which the Data Warehouse will be operative. Security levels, roles and organizational compartments can be defined according to the needs of the organization [RODOLFO et al. 2006]. As creating new profile include creating new stereotypes, [RODOLFO et al. 2006] defined a package that includes all the stereotypes that will be necessary in the profile. This profile contains four types of stereotypes [RODOLFO et al. 2006]:

1) Secure class and secure Data Warehouses stereotypes (and stereotypes inheriting information from them) that contain tagged values associated to attributes (model or class attributes), security levels, user roles and organizational compartments.
2) Attribute stereotypes (and stereotypes inheriting information from attributes) and instances, which have tagged values associated to security levels, user roles and organizational compartments.
3) Stereotypes that allow us to represent security constraints, authorization rules and audit rules.
4) UserProfile stereotype, which is necessary to specify constraints depending on particular information of a user or a group of users.

In Figure 21, we can see the tagged values associated to each one of the stereotypes. For example, 'SecureDW' stereotype has the following values associated: Classes, SecurityLevels, SecurityRoles and SecurityCompartments. The tagged values they have defined are applied to certain components that are especially particular to MD modeling, allowing DW designers to represent them in the same model and in the same diagrams that describe the rest of the system [RODOLFO et al. 2006]. In Table 2, the necessary tagged values in the profile are shown. These tagged values will represent the sensitivity information of the different elements of the MD modeling (fact class, dimension class, base class, etc.), and they will allow to specify security constraints depending on this security information and on the value of attributes of the model [RODOLFO et al. 2006].

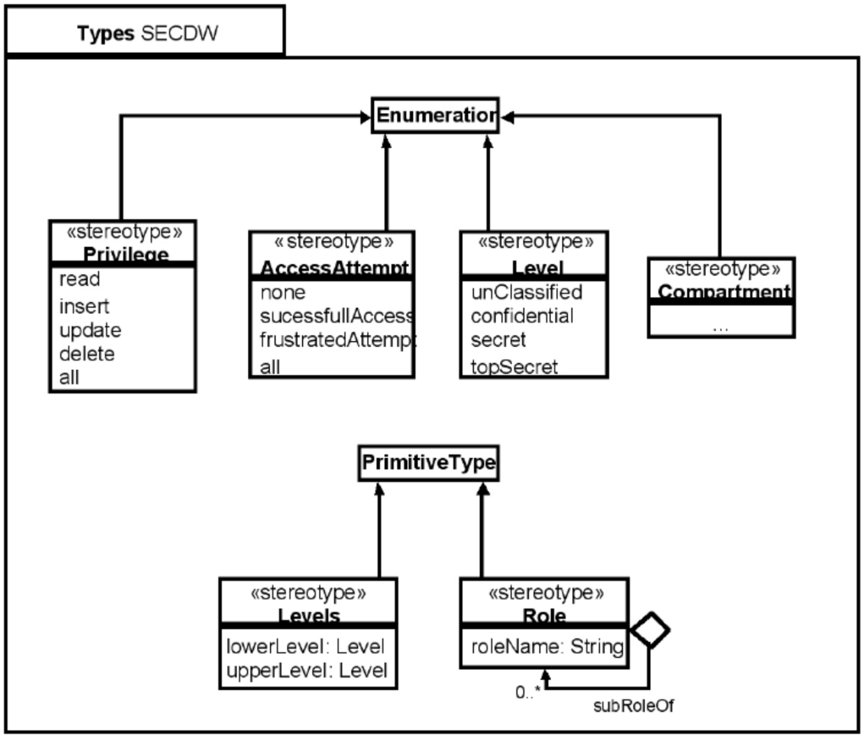


Figure 20. Values associated to new data types [RODOLFO et al. 2006].

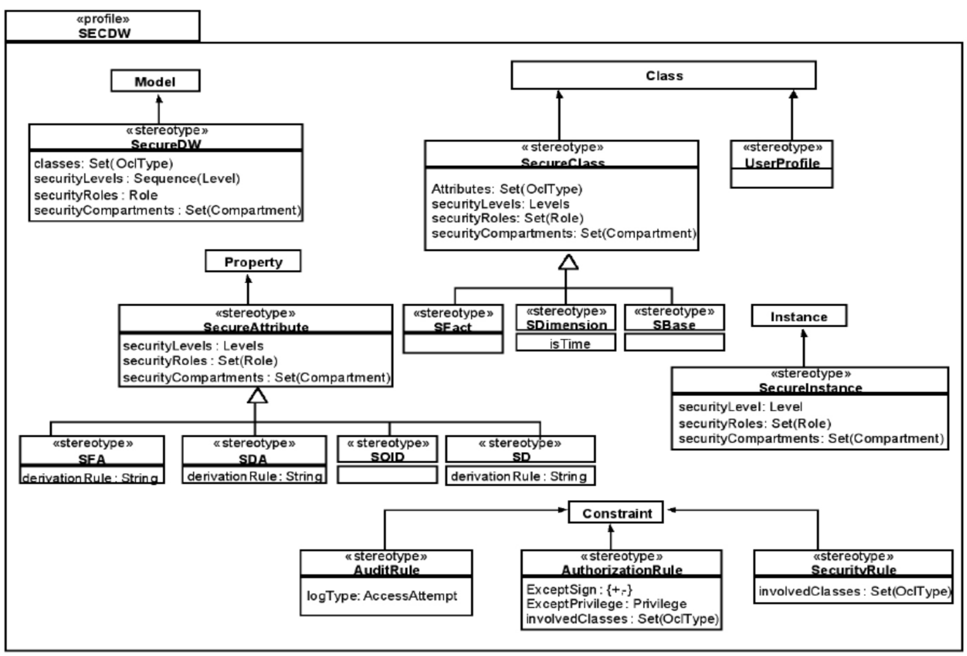


Figure 21. New stereotypes [RODOLFO et al. 2006].

[EDUARDO et al. 2006a] proposed an approach for developing secure Data Warehouses with a UML extension. They proposed an Access Control and Audit (ACA) model for DWs by specifying security rules in the conceptual MD modeling. They also defined authorization rules for users and objects and assigned sensitive information rules and authorization rules to the main elements of a MD model. Moreover, they specified certain audit rules allowing analyzing user behaviors [EDUARDO et al. 2006a].

Table 2. Tagged values [RODOLFO et al. 2006].

Name	Type	Description	Default Value
Classes	Set(Ocltype)	It specifies all classes of the model. This new tagged value is useful in order to navigate through all classes of the model.	Empty set
Attributes	Set(OclType)	It specifies all attributes of the class. This new tagged value is useful in order to navigate through all attributes of the model.	Empty set
Security-Levels	Levels	It specifies the interval of possible security level values that an instance of this class can receive.	The lowest level (if we consider traditional levels, should be 'Unclassified')
Security-Roles	Set(Role)	It specifies a set of user roles. Each role is the root of a subtree of the general user role hierarchy defined for the organization.	The set composed of one role that is the role hierarchy defined for the model
Security-Compartments	Set (Compartment)	It specifies a set of compartments. All instances of this class can have the same user compartments, or a subset of them.	Empty set of compartments
Name	Type	Description	Default Value
LogType	AccessAttempt	It specifies whether the access has to be recorded: none, all access, only frustrated accesses, or only successful accesses.	None
Involved-Classes	Set(OclType)	It specifies the classes that have to be involved in a query to be enforced in an exception.	Empty
ExceptSign	{+,-}	It specifies if an exception permits (+) or denies (-) access to instances of this class to a user or a group of users.	+
Except-Privilege	Set(Privilege)	It specifies the privileges the user can receive or remove.	Read
isTime	Boolean	It indicates whether dimension represents a time dimension or not.	False
derivationRule	String	If the attribute is derived, this tagged value represents the derivation rule.	Empty

Figure 22 depicts an example of MD model after applying the security information and constraints proposed by [EDUARDO et al. 2006a]. [EDUARDO et al. 2006a] asserted the importance of specifying security aspects from the early stages of the Data Warehouses (DWs) and enforce them. Because Data Warehouses (DWs), MultiDimensional (MD) Databases, and On-Line Analytical Processing(OLAP) Applications have been used as a very powerful mechanism for discovering crucial business information, it is important to handle security aspects of DWs in proper way [EDUARDO et al. 2006a]. Considering the extreme importance of the information managed by DWs and OLAPs, it has been essential to specify security measures from the early stages of the DW design in the MD modeling process, and enforce them. [EDUARDO et al. 2006a] considered security issues of as an important element in DW model that has not been handled at conceptual level, so they specified confidentiality constraints to be enforced and modeled during developing the conceptual MD model of DW [EDUARDO et al. 2006a].

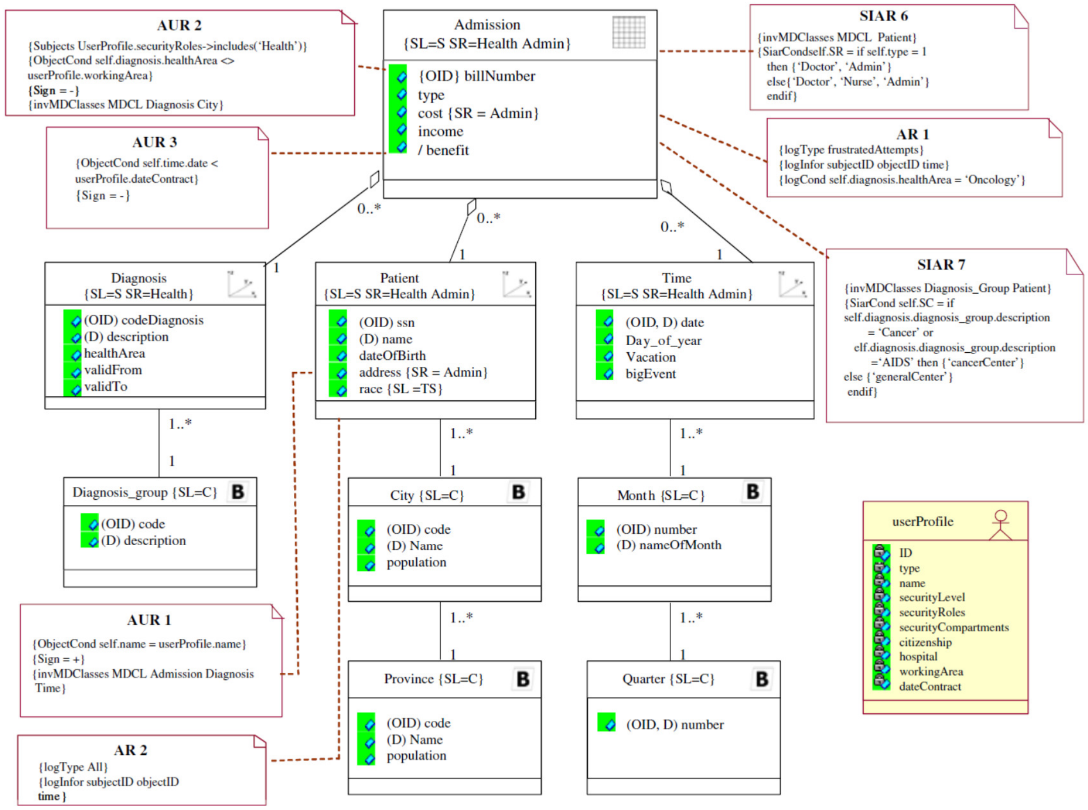


Figure 22. Example of MD model with security information and constraints [EDUARDO et al. 2006a].

One key advantage of their approach was that they accomplished the conceptual modeling of secure DWs independently of the target platform where the DW had to be implemented, allowing the implementation of the corresponding DWs on any secure commercial database management system. Finally, they presented a case study to show how a conceptual model designed with their approach could be directly implemented on top of Oracle 10g [EDUARDO et al. 2006a].

[EMILIO et al. 2007] presented a framework for the development of secure Data Warehouses (DWs) based on MDA and QVT. This framework was a natural continuation of their previous work [JOSE-NORBERTO et al. 2005] and of the results presented in [EMILIO et al. 2006], [RODOLFO et al. 2006], [EDUARDO et al. 2006b]. They proposed a framework based on Model-Driven Architecture (MDA) for the development of secure Data Warehouses that covers all the phases of design (conceptual, logical and physical) and embedded security measures in all of them. Moreover, transformations between models were clearly and formally executed by using Query-View-Transformation [OMG 2008], to obtain a traceability of the security rules from the early stages of development to the final implementation [EMILIO et al. 2007].

As stated by [EMILIO et al. 2007], [JOHN et al. 2003] proposed an access and audit control model integrated with a Unified Modeling Language extension, this allowing the development of secure MultiDimensional models at conceptual level. This proposal was promising, but still it did not cover all the stages of a DW development cycle [EMILIO et al. 2007]. [EMILIO et al. 2007] mentioned that current specialized literature comprises several proposals to integrate security with the MDA technology ([CAROL et al. 2003], [DAVID et al. 2006], [SIVANANDAM et al. 2004], [LANG et al. 2004]), but all of them are related with information systems, access control, security services and secure distributed applications, so none of them, is related with the design of secure DWs [EMILIO et al. 2007].

Main objective of [EMILIO et al. 2007] was the proposal of an architecture that transforms security requirements from the conceptual level up to the logical level. They defined QVT relations that allow DW designers to represent at logical level all security and audit requirements captured at the stage of conceptual modeling of the secure Data Warehouses. The application of QVT transformation rules to the security MultiDimensional (SMD) PIM allowed the development of different SMD PSMs, thus facilitating the representation at a logical level of all security and audit requirements captured at earlier stages of DW design. Afterwards, each SMD PSM was directly converted into code [EMILIO et al. 2007].

The main contributions of [EMILIO et al. 2007] work are: the development of DWs was reduced to creating an SMD PIM and its corresponding QVT relations, the time and effort invested in the development of DWs were shortened, the transition between different models and the final implementation was guaranteed, they reached interoperability, portability, adaptability and reusability by employing MDA technology [EMILIO et al. 2007].

Figure 23 shows the Secure MultiDimensional MDA architecture for the development of secure Data Warehouses. The upper section of Figure 23 shows the CIM that declares the requirements for the DW. It represents a perception on the DW within its business environment, so it plays an important role in reducing the gap between business people and those who are experts in the design and development of the DW which needs to satisfy the requirements [EMILIO et al. 2007]. By means of the transformation T1, the Secure MultiDimensional PIM can be obtained, which is located at conceptual level. The T2 transformation derives the Secure MultiDimensional PSM (SMD PSM). This transformation is not unique, as other secure PSMs are possible [EMILIO et al. 2007 a]. [EMILIO et al. 2007] used the security UML profile, SECDW, developed by [RODOLFO et al. 2006] to represent the main security requirements for the conceptual modeling of the DW. Figure 24 represents the SECDW metamodel used in the design of SMD PIM [EMILIO et al. 2007 a].

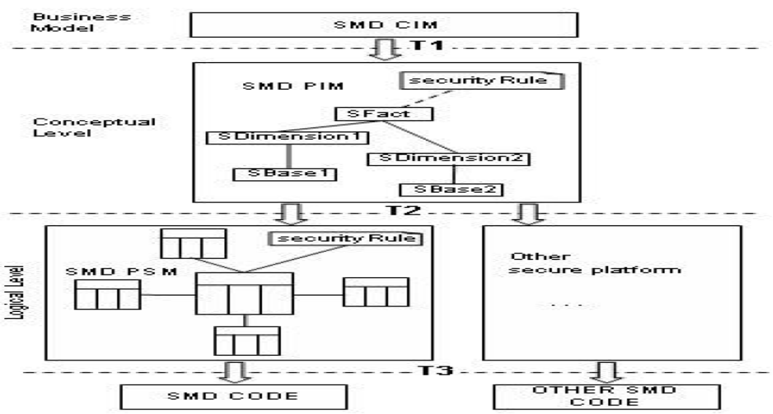


Figure 23. A framework for the development of secure Data Warehouses [EMILIO et al. 2007a].

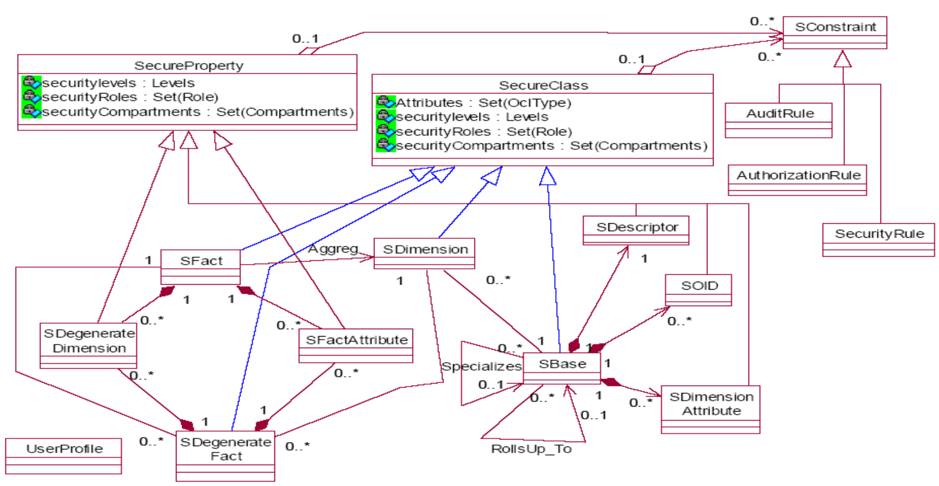


Figure 24. The SECDW metamodel used in the design of SMD PIM [EMILIO et al. 2007].

Figure 25 presents the SECRDW metamodel that was designated as PSM. In order to distinguish the security aspects it comprises, it will be called the Secure MultiDimensional PSM (SMD PSM). This metamodel allows DW designers to represent Schema, Tables, Columns, Primary, Foreign keys and the needed security aspects of the system [EMILIO et al. 2007]. Figure 26 represents the textual notation for the main QVT transformations, i.e., the SMD PIM to SMD PSM transformation [EMILIO et al. 2007].

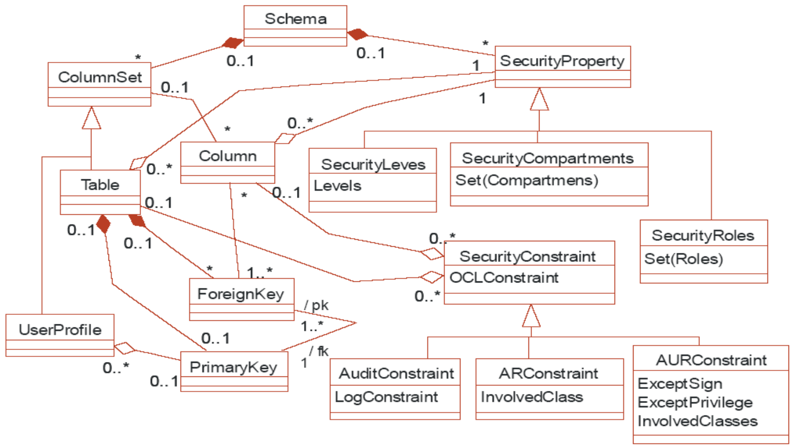


Figure 25. The SECDW metamodel used in the design of the SMD PSM [EMILIO et al. 2007].

```
Transformation SMD To SREL(SMD: SECDW)
     SREL: SECRDW)
{key Table{name, Schema};
key Column {name, owner};
key UserProfile{name, Schema};
key PrimaryKey{name, owner};
key ForeignKey{name, owner};
key SecurityProperty{name, owner};
key SecurityConstraint(name, owner);
top relation SecureDW2Schema{}
top relation UserProfile2RUserProfile{}
top relation SFact2Table {}
top relation SDegenerateFact2Table{}
top relation SDimension2Table{}
  //Association SFact with SDimension
top relation AssocSF_D2FKey{}
  //Association SDegenerateFact with SDimension
top relation Assoc SDF_SD2FKeyFKey
// Association SDegenerateFact with SFact
top relation AssocSDF_SF2FKey{}}
```

Figure 26. Textual notation for the SMDPIM to SMDPSM transformation [EMILIO et al. 2007].

[EMILIO et al. 2008a] presented comprehensive requirement analysis approach for considering security in early stages of DW development life cycle. In this paper, they focus on describing a comprehensive requirement analysis approach for DWs that comprises two parts. The first one is Functional Requirement analysis and the second one is QoS requirement analysis. Requirement analysis approaches for DWs have focused attention merely on information needs of top management and decision makers, without taking into consideration other kinds of QoS requirements such as performance or security [EMILIO et al. 2008a].

Modeling these requirements in the early stages of the development is a foundation stone for building a DW that satisfies user wants and needs. [EMILIO et al. 2008a] specified the two kinds of requirements for data warehousing as QoS requirements and Functional Requirements and jointed them in a broad approach based on MDA. This permitted a separation of concerns to model requirements without losing the connection between Functional Requirements and quality-of-service requirements [EMILIO et al. 2008a]. Finally, [EMILIO et al. 2008a] introduced a security requirement model for data warehousing, and a three-step process for modeling security requirements.

Based on [EMILIO et al. 2008a], the development of a DW should focus on the design of a conceptual MD model. As shown in Figure 27, the specification of this model must be driven by an analysis of operational data sources, Functional Requirements, and QoS requirements. By this way the design of conceptual MD model satisfies user expectations and agrees with the operational sources.

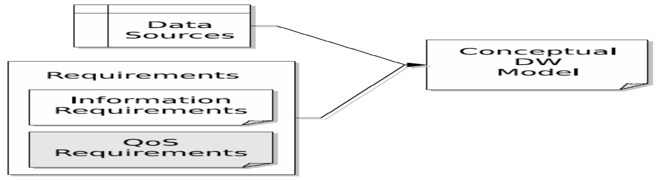


Figure 27. QoS Requirements are needed as input for Data Warehouse design [EMILIO et al. 2008a].

For modeling the Functional Requirements, [EMILIO et al. 2008a] used a UML profile for the i* modeling framework [YU 1997]. As depicted in Figure 28, the i* modeling framework provides mechanisms to represent different DW actors, their dependencies, and for structuring the business goals that the organization wants to achieve with the DW. Two models are used in i*: the strategic dependency (SD) model for describing the dependency relationships among various actors in an organizational context, and the strategic rationale (SR) model, used to describe actor interests and concerns, and how they might be addressed [EMILIO et al. 2008a].

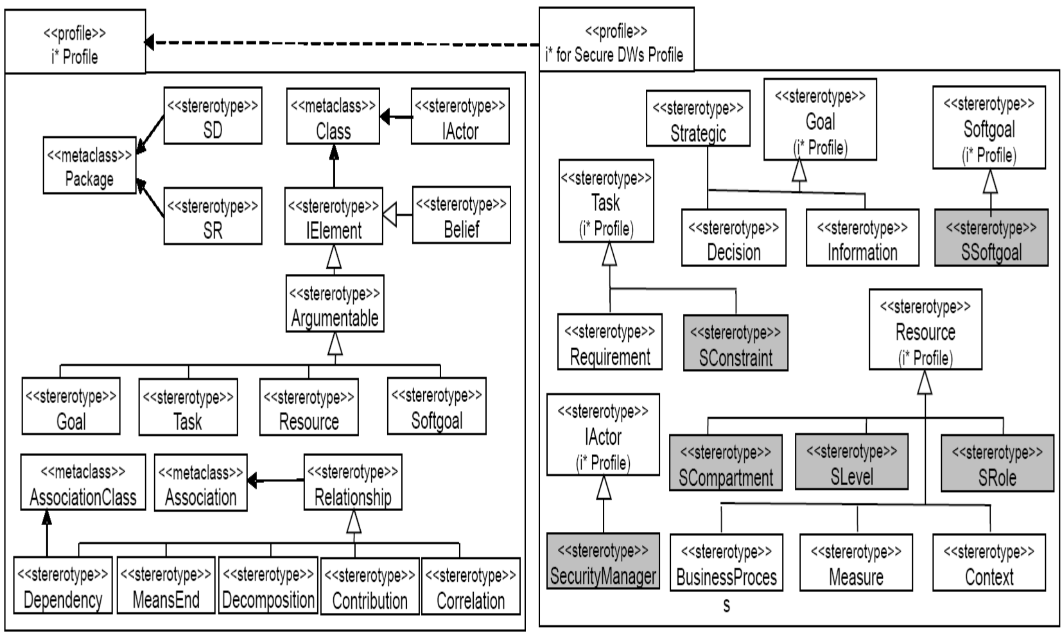


Figure 28. Overview of the profiles for i* modeling in the DW domain. [EMILIO et al. 2008a].

As stated by [EMILIO et al. 2008a], once the Functional Requirements have been identified, the model is improved by adding QoS requirements. QoS requirements are diverse. To not overlook any important feature, it is mandatory to use a framework of QoS requirements in DW. Figure 29 depicts a framework for capturing the many different aspects that must be considered when designing a DW. The Figure is based on the type catalogue for Non-Functional Requirements for DW design introduced by [PAIM et al. 2002].

Discussion:

All previous works ([RODOLFO et al. 2006], [EDUARDO et al. 2006a], [EMILIO et al. 2007], [EMILIO et al. 2008a]) covered all security aspects that should be handled at conceptual level of DW project. Hence this thesis considers the security requirements of traditional Data Warehouses as a subject that has been addressed in proper way and it is in a fully mature status. However, these works addressed one part of QoS “Non-Functional Requirement”; other parts of QoS such as DW performance still not handled at conceptual level using MDA standard. Moreover, these works did not address data sources analysis and we encourage reader to refer to [JOSE-NORBERTO et al. 2007a], [JOSE-NORBERTO et al. 2007c] for a wider explanation about operational data sources analysis.

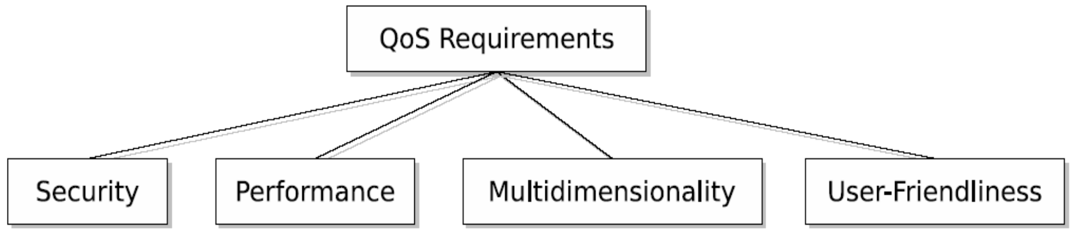


Figure 29. Issues to be considered during Data Warehouse design [EMILIO et al. 2008a].

3.3.1 Designing Spatial Data Warehouse using MDA Techniques

[OCTAVIO et al. 2008] presented an MDA approach for spatial Data Warehouse development. In this paper, they presented a spatial extension for the MD model to embed spatiality on it. Then, they formally defined a set of Query-View-Transformation rules which allow obtaining a logical representation in an automatic way. Finally, they showed how to implement the MDA approach in their Eclipse-based tool. [OCTAVIO et al. 2008] presented set of limitations regarding development of spatial Data Warehouse and how their new approach solves such limitations. They also mentioned that several conceptual approaches have been proposed for the specification of the main MultiDimensional (MD) properties of the spatial Data Warehouses (SDW) [BIMONTE et al. 2005],[ELZBIETA et al. 2004b]. However, these approaches often fail in providing mechanisms to automatically derive a logical representation and the development time and cost is increased. Furthermore, the spatial data often generates complex hierarchies (i.e., many-to-many) that have to be mapped to large and non-intuitive logical structures (i.e., bridge tables) [OCTAVIO et al. 2008].

In their work, the authors introduced spatial data on their previous work [JOSE-NORBERTO et al. 2005], [JOSE-NORBERTO et al. 2008] to accomplish the development of SDWs using MDA. Therefore, they have focused on (i) extending the conceptual level with spatial elements, (ii) defining the main MDA artifacts for modeling spatial data on a MD view, (iii) formally establishing a set of QVT transformation rules to automatically obtain a logical representation tailored to a relational spatial database (SDB) technology, and (iv) applying the defined QVT transformation rules by using their MDA tool, thus obtaining the final implementation of the SDW in a specific SDB technology (PostgreSQL [POSTGRESQL 2009] with the spatial extension PostGIS [POSTGIS 2009]) [OCTAVIO et al. 2008]. Figure 30 shows a symbolic diagram of the proposed approach: from the PIM (spatial MD model), several PSMs (logical representations) can be obtained by applying several QVT transformations [OCTAVIO et al. 2008].

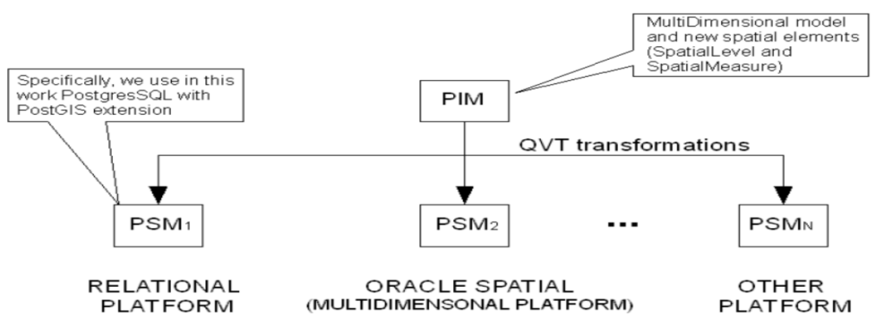


Figure 30. Overview of MDA approach for MD modeling of SDW repository [OCTAVIO et al. 2008].

Based on their previous profile [SERGIO et al. 2006], they enriched the MD model with the minimum required description for the correct integration of spatial data, coming in spatial levels and spatial measures. Then, they implemented these spatial elements in the base MD UML profile [OCTAVIO et al. 2008]. Finally, they added a property to these new stereotypes in order to geometrically describe them. All the allowed geometric primitives use to describe elements are group in an enumeration element named GeometricTypes. These primitives are included on ISO [ISO 2009] and OGC [OGC 2009] SQL spatial standards, in this way they ensured the final mapping from PSM to platform code. The complete profile can be seen on Figure 31 [OCTAVIO et al. 2008].

Discussion:

[OCTAVIO et al. 2008] proposal is an MDA-oriented framework for the development of DWs to integrate spatial data and develops Spatial DWs. They used spatial MD modeling profile as a PIM and the CWM relational package as a PSM; so there is a space in the research area to handle other PSMs such as objects and Object relational platforms. Moreover, [OCTAVIO et al. 2008] did not handle QoS aspects for spatial DW which are a hot topic for joining MDA with SDW.

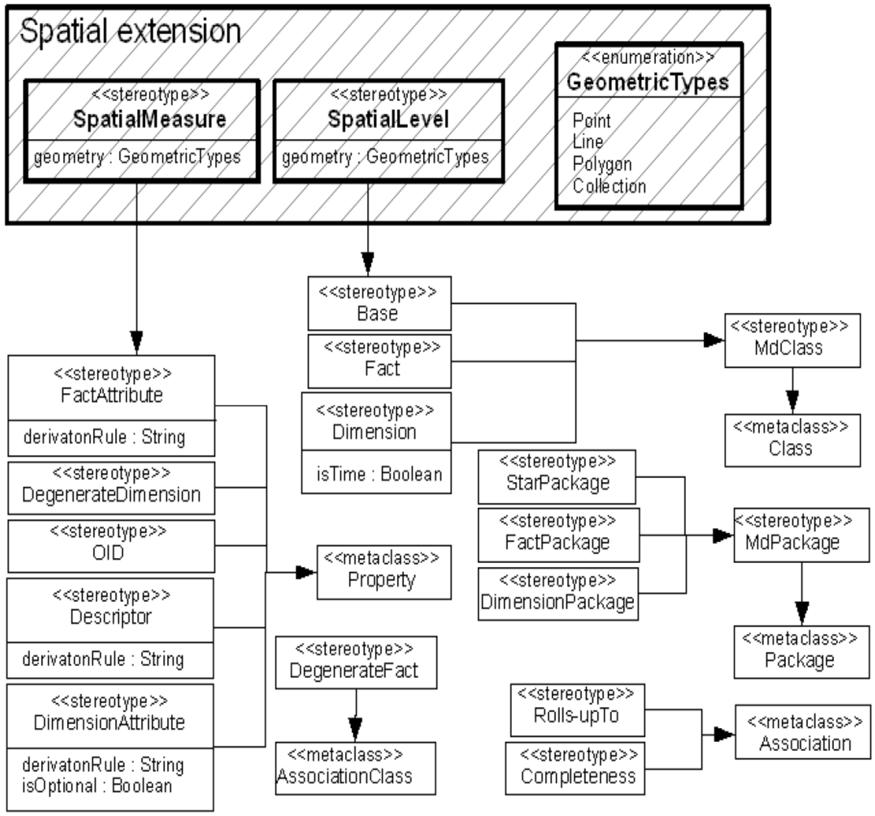


Figure 31. UML Profile for conceptual MD modeling with spatial elements in order to support spatial data integration [OCTAVIO et al. 2008].

3.3.2 New Business-level Security UML Profile

[JUAN et al. 2009a] proposed a profile which uses the Unified Modeling Language (UML) extensibility mechanisms. This profile allows us to define security requirements for DWs at the business level, taking into account the Functional Requirements modeled with their previous profile [JOSE-NORBERTO et al. 2007b]. The proposal is aligned with Model-Driven Architecture (MDA), thus permitting the transformation of security requirements throughout the entire DW life cycle [JUAN et al. 2009a]. Finally, in order to show the benefits of the proposed profile, they develop a case study related to the management of a pharmacy consortium business.

In Figure 32, [JUAN et al. 2009a] tried to present the extensions proposed in order to accommodate DW development to the MDA approach. The CIM is based on an extension of the i* framework [YU 1997] proposed in [JOSE-NORBERTO et al. 2007b], which deals solely with Functional Requirements for DW design at the business level. The PIM corresponds with an extension of the Unified Modeling Language (UML) profile presented in [RODOLFO et al. 2006], which reuses the results of [SERGIO et al. 2006]. This profile considers the main properties of secure MD modeling at the conceptual level. The PSM corresponds with an extension of the Common Warehouse Metamodel (CWM) at the logical level [EMILIO et al. 2008b], and Code with implementation at the physical level (DBMS level).

LEVELS	MDA	DWs DESIGN	EXTENSION
Business	CIM	Requirements Analysis	i* metamodel
Conceptual	PIM	Multidimensional Secure Model	UML metamodel
Logical	PSM_1 ··· PSM_n	Relational Secure Model	The Relational Package from CWM metamodel
Physical	$Code_1$ ··· $Code_n$	SGBD implementation	None

Figure 32. Aligning the design of secure DWs with MDA [JUAN et al. 2009a].

[JUAN et al. 2009a] used the metamodels presented in [RODOLFO et al. 2006] and [EMILIO et al. 2008b] in order to define QVT relations to transform PIM into PSM for secure DW design. This set of QVT relations has been validated through the development of a case study [JUAN et al. 2009a]. In a previous work [JOSE-NORBERTO et al. 2007b], [JUAN et al. 2009a] have employed i* modeling and the MDA framework in order to model goals and functional requirements for DWs. This proposal defines a UML profile based on i* modeling for the DW design, which allows to formalize i* diagrams in order to model a CIM. However, the approach focuses only on Functional Requirements; it does not include security as a special Non-Functional Requirement type. Therefore, [JUAN et al. 2009a] proposed a new UML profile that reuses the above mentioned profile [JOSE-NORBERTO et al. 2007b], whilst adding security requirements for DWs at the business level.

The work provided by [JUAN et al. 2009a] allows an integrated method to develop a complete methodology to build secure DWs. The main benefits of their proposal are: (i) the adaptation of the i* framework to define and integrate both security and information as Functional Requirements into a secure CIM for DWs, (ii) a guarantee of consistency since the profile avoids the situation of having different definitions and properties for the same concept throughout a model, and (iii) an attempt to create a proposal which is more understandable to both DW designers and final users [JUAN et al. 2009a].

Discussion:

This effort conducted by [JUAN et al. 2009a] represents a fully integrated MDA framework for designing secure DW. This work is the fruitful result of many years of valuable efforts on MDA and DWs. However applying the same approach on temporal and spatial DWs and working on other types of QoS (e.g., performance) will be an added value to the literature.

3.3.3 Conceptual OLAP Platform-Independent Queries

[JESUS et al. 2008] presented a proposal that bridges the semantic gap between the DW conceptual and logical models. The development of Data Warehouses is based on specifying both the static and dynamic properties of on-line analytical processing (OLAP) applications by means of the conceptual model. Then, developers design its logical counterpart where platform-specific details such as performance or storage are also considered. However, it is well known the existence of a semantic gap between the conceptual and logical levels that decreases the feasibility of their mapping [JESUS et al. 2008].

In order to bridge this gap, [JESUS et al. 2008] proposed the use of conceptual OLAP queries, i.e., platform independent, which can be automatically traced to their logical form in a coherent and integrated way. The proposed solution provided by [JESUS et al. 2008] comprises (i) the definition of an OLAP algebra that response analysts' information needs at the conceptual level, and (ii) a model-transformation architecture that automatically manages and derives from them the different logical designs, being aware of the platform-specific details.

This approach takes advantage of UML, OCL, or QVT, that enable an integrated solution for querying Data Warehouses. Its feasibility has been shown by specifying in OCL each of the most common OLAP operation in every OLAP algebra [ROMERO et al. 2007]. OCL has been successfully employed in order to automatically derive both data structures and OLAP queries for the well-known star and snowflake logical schema in a SQL-based relational database. Based on [JESUS et al. 2008], the main benefits of this approach are: it is the first approach that employs OCL for specifying OLAP data cubes using textual format. Second, since it is difficult to intuitively represent the dynamic part of a MultiDimensional model within a pure mathematical notation, the proposed approach provides an intuitive and easy way to model queries integrated into the conceptual data modeling. Third, querying at the conceptual level implies that analysts do not need to be aware of which design decisions have been taken in order to implement the conceptual model to be queried [JESUS et al. 2008]. Figure 33 shows the big picture of the proposed MDA platform-independent queries.

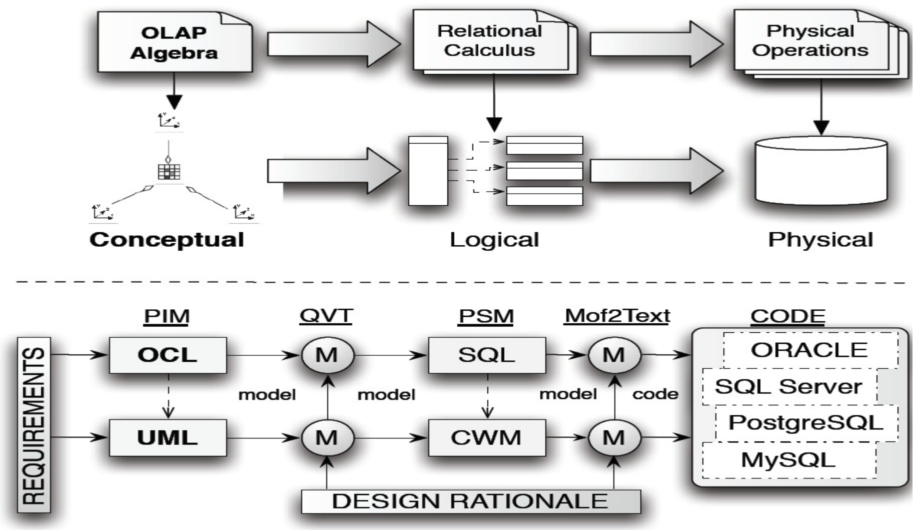


Figure 33. MDA for platform-independent queries [JESUS et al. 2008].

Discussion:

This work has addressed the existence of a semantic gap between the conceptual and logical levels that decreases the feasibility of their mapping. Relational databases are the platform that has used in this work. This opens new research area for implementing similar efforts on different platforms such as object and object relational databases.

3.3.4 MDA Framework for Designing Spatial DWs

[OCTAVIO et al. 2009] presented data model for representing and querying geographic information customized on MultiDimensional structure of the data and the OLAP analysis technique. Thus, [OCTAVIO et al. 2009] have defined formally with OCL and modeled with UML profiles the customization and the DW layers as presented in Figure 34. This framework, which is based on the previous work [OCTAVIO et al. 2008], addresses the design of the whole Geographical Data Warehouse (GDW) system by align every component with the different MDA viewpoint.

[OCTAVIO et al. 2009] could use the MDA and QVT techniques to build a spatial conceptual model and transform it into code. By this effort, the development of GDWs is simplified in just two tasks: (i) the development of conceptual models for each component; and (ii) the development of the corresponding QVT transformations to automatically generate the GDW implementation from every conceptual model developed [OCTAVIO et al. 2009].

This solution helps developers to directly include spatial data at conceptual level, while decision makers can also conceptually query them without being aware of logical details. [OCTAVIO et al. 2009] also presented a practical application in order to show the benefits of their proposal. They have implemented their methodology on Eclipse platform by using plugin extensions. With this developing tool they derivate a GIS OLAP application example by building conceptual models. This application is also implemented in Eclipse platform and also uses Mondrian as OLAP Server and uDig as map interface (a GIS framework for Eclipse). [OCTAVIO et al. 2009] also have used PostgreSQL with the spatial extension PostGIS for data implementation.

Discussion:

The work done by [OCTAVIO et al. 2009] is similar to the previous work done by [JOSE-NORBERTO et al. 2005] in which an MDA framework for traditional DWs has been created; but this work applied MDA techniques to a spatial DWs. However, this work opens a set of GDW research problems such as developing secure GDW profile, applying MDA concepts to a spatial data mining and what-if-analysis application, and using MDA for building GDW ETL programs.

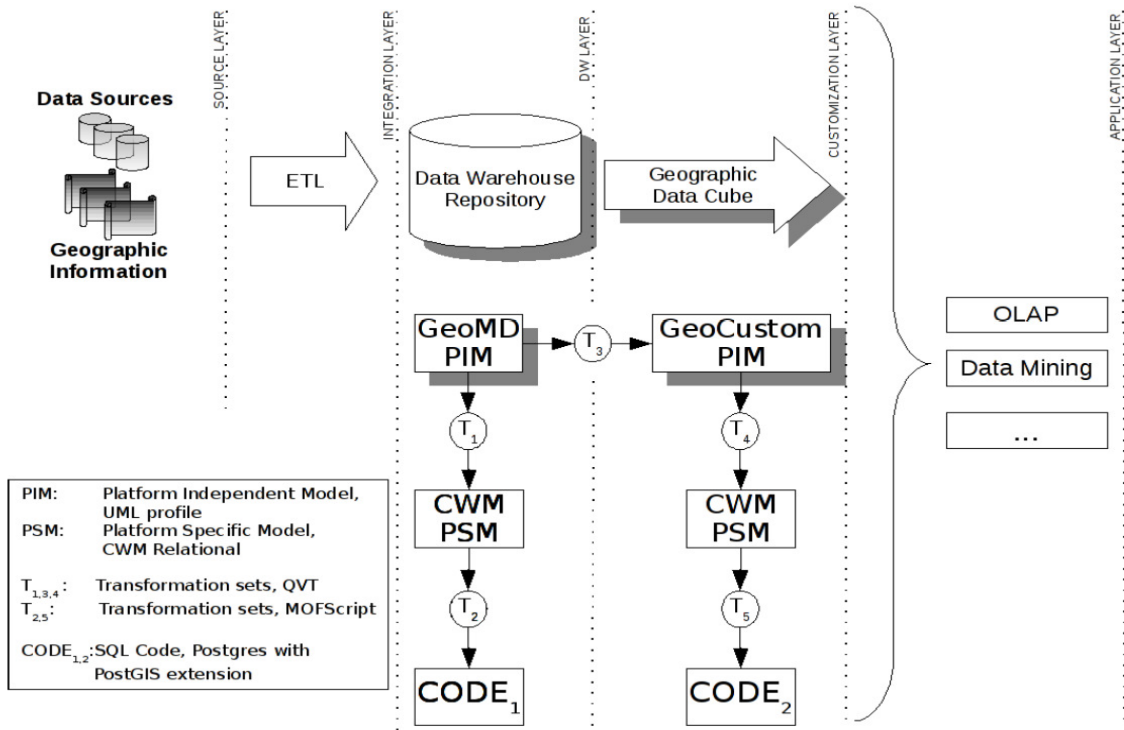


Figure 34. Model driven development framework able to integrate geographic capabilities in multilayered spatial Data Warehouses [OCTAVIO et al. 2009].

3.4 MDA Secure Engineering Process for DWs

[JUAN et al. 2009b] proposed a secure engineering process for DWs, by eliciting and developing both functional and security aspects as Non-Functional Requirements at the business level. They called this methodology a Secure Engineering process for Data Warehouses (SEDAWA) which composed of four phases comprising of several activities and steps, and five disciplines which cover the whole DW design [JUAN et al. 2009b]. This methodology can be summarized as follows. First a secure CIM is built by using the three activities supported by an adjustment of the i* framework. Second, the secure CIM is transformed and developed by using QVT transformations throughout the DW life cycle. The proposed methodology is MOF-compliant as a result of the application of Software Process Engineering Metamodel Specification (SPEM), i.e., according to the four layer architecture from OMG, it belongs to the M1 layer [JUAN et al. 2009b].

The greatest contribution of [JUAN et al. 2009b]'s work is that all the security and audit requirements elicited during the early phases are modeled, developed and defined throughout the entire DW life cycle. Therefore, both the time and effort invested in the development of DWs are narrowed, the transition between different models and the final implementation is guaranteed, and that it is possible to achieve interoperability, portability, adaptability and reusability by utilizing MDA technology [JUAN et al. 2009b].

SPEM is a process metamodel used to describe a concrete software development process or a family of related software development process. The SPEM specification is structured as a UML profile, and provides a complete MOF-based metamodel [OMG 2005A].

The SPEM metamodel offers the constructs and semantics required for the software development process, which need the use of Unified Modeling Language (UML). The SPEM stand-alone metamodel is built by extending a subset of the UML metamodel [JUAN et al. 2009b]. Figure 35 shows part of the SPEM metamodel that has been used in the proposed engineering process, which is supported by the Core and ProcessComponent packages [JUAN et al. 2009b].

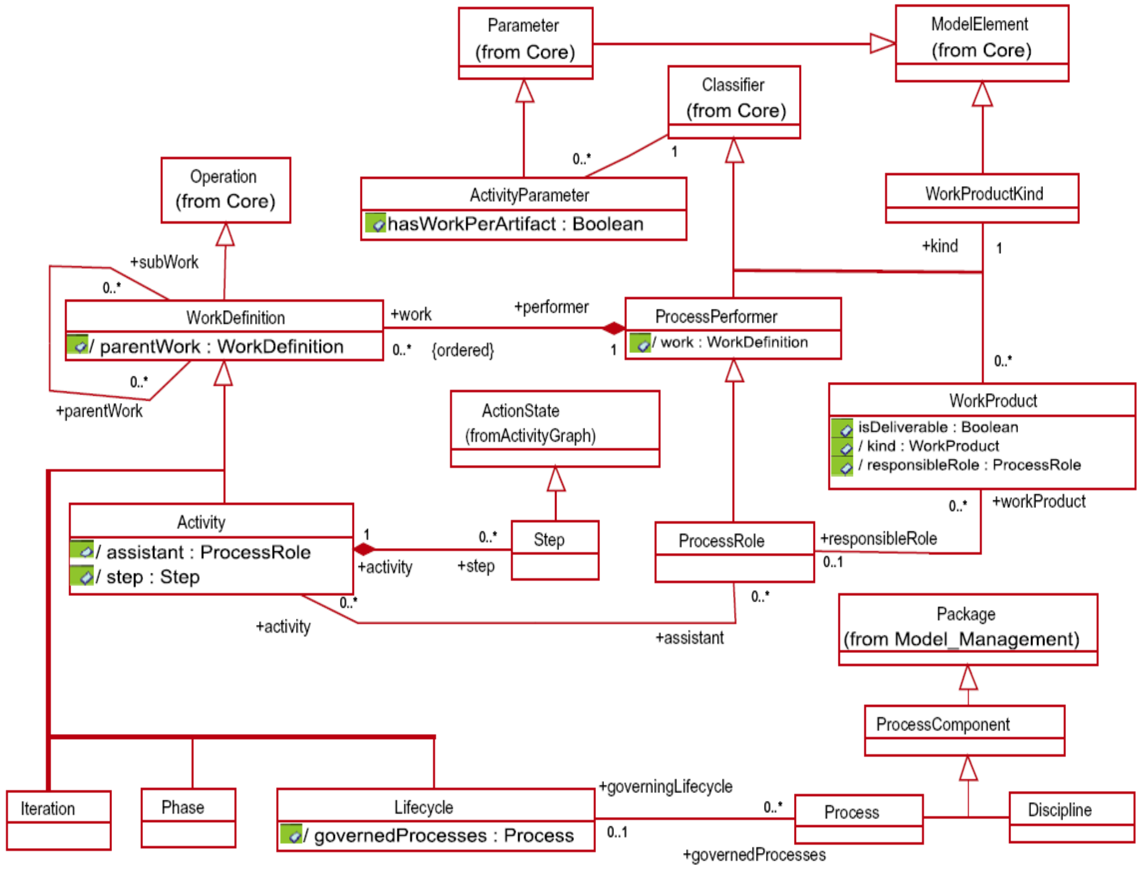


Figure 35. Fragment of SPEM metamodel employed. [JUAN et al. 2009b].

As presented in Figure 36, SEDAWA is structured into four consecutive phases: elicitation, modeling, implementation, and test–delivery. The iterative style has been applied to the phases of SEDAWA methodology. Five disciplines have been defined: requirements analysis, conceptual design, logical design, physical design and post-development review [JUAN et al. 2009b]. The engineering process begins with the Enterprise Architecture WorkProduct as input for activity A1.1. The Enterprise Architecture contains designs of the business processes, organizational structures, components, physical resources, products and services from the organization. This WorkProduct can be used, by applying activities A1.1, A1.2 and A1.3 from the Elicitation phase to obtain three models: (1) GOModel which contains informational requirements for DWs; (2) SOModel which contains security requirements for DWs; and (3) GSAModel which merges the above models and constitutes a secure CIM for DWs [JUAN et al. 2009b].

The Modeling phase is conducted by activities A2.1, A2.2. Activity A2.1 receives as input the WorkProducts GSAModel and Secure MD metamodel. Activity A2.2 receives as input the MD model WorkProduct obtained from activity A2.1 and the operational sources WorkProduct that will serve to populate the secure DWs repository. The implementation phase is executed out by activity A3.1, which accepts as input the enriched secure MD model Work- Product obtained from activity A2.2. In addition A3.1 accepts the SECure Relational Data Warehouses (SECRDW) metamodel and the DBMS specific WorkProducts. Finally, the test–delivery phase encloses the activity A4.1 in order to validate, test and deliver the secure DW repository [JUAN et al. 2009b].

Discussion:

[JUAN et al. 2009b]'s work represents the first effort for having a process-oriented MDA-based DW development methodology. This effort opens a new research areas regarding developing Capability

Maturity Model integration (CMMI) [DENNIS et al. 2008], International Standard Organization (ISO) or Project Management Institute (PMI) DW development methods that are an MDA-based once. This new type of engineering processes will add and join the needed project management skills (i.e. project initiation, project planning, Configuration management, etc.) to the method to be integrated with the MDA techniques.

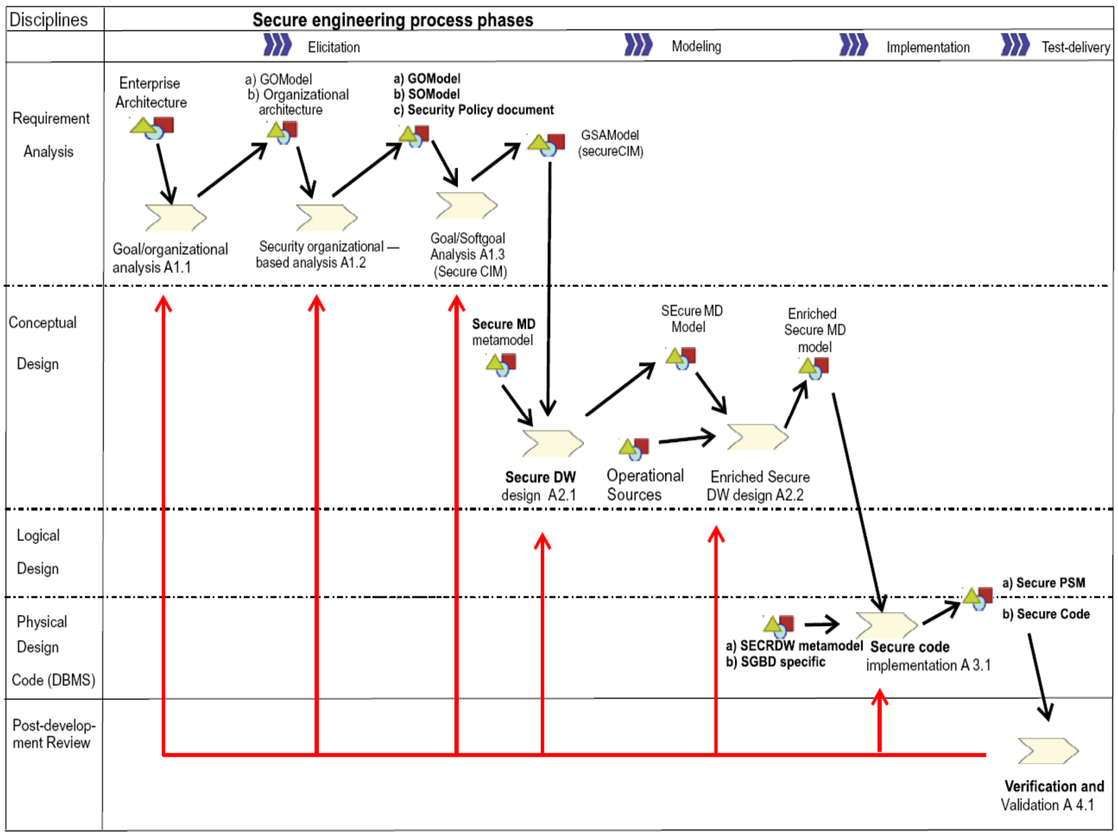


Figure 36. Secure engineering process for DWs. [JUAN et al. 2009b].

4. OPEN RESEARCH PROBLEMS

After surveying and highlighting the current research direction in MDA and DW, a set of open research problems has been identified. First, most of traditional Data Warehouse development frameworks (non MDA ones) such as ([WINTER et al. 2003], [PAOLO et al. 2005], [NAVEEN et al. 2004]) focused only on Functional Requirement gathering to build the conceptual MetaDimensional model (MD) that identify the main element of this MD such as Fact tables and Dimension tables. These approaches did not address the QoS requirement (Non-Functional Requirement) such as security and performance that the end user expects to have. Other approaches such as [PAIM et al. 2002] tried to address QoS (Non-Functional Requirements) for DW but in an isolated way from Functional Requirement analysis and without applying MDA concepts and techniques.

As stated earlier, [EMILIO et al. 2008a] presented a work that was an MDA-based for Data Warehouse development framework that focused on Functional and Non-Functional Requirements (QoS). [EMILIO et al. 2008a] focused on one element of QoS which is the security requirements, leaving a space for handling other QoS requirement specifications such DW performance and user friendliness requirements. [DANIAL et al. 2001] presented a tool-based solution for metamodel exchange problem by using a CWM-based tool which facilitates the process of metamodel transfer between different environments. [DANIAL et al. 2001] left a space for enhancing the tool to support other CWM specifications. This work focused only on the relational part of CWM called CWM Relational while CWM is a specification for modeling metadata for relational, non-relational, MetaDimensional systems.

[JOSE-NORBERTO et al. 2005], as stated earlier, presented an overall MDA approach for all DW design stages. However, [JOSE-NORBERTO et al. 2005]'s work was in very high-level format which hide behind it a set of open research areas such as using MDA for other layers of DW, such as applying MDA in integration layer(ETL application), Application layer (OLAP , Data Mining, what-if-analysis application stage and in customization layer (Data Cubes).

As mentioned previously, [JUAN et al. 2009a] has presented a UML 2.0 profile which allows DW designer to define the security requirements for DWs at the business level. However, the definition and the implementation of the QVT relations in order to establish a transformation between the CIM and the PIM levels have not been addressed. [OCTAVIO et al. 2009], as stated earlier, presented an MDA Framework for Designing Spatial DWs. This work opens a new set of GDW research problems such as developing secure GDW profile, applying MDA concepts to a spatial data mining application and using MDA for building GDW ETL programs. [JUAN et al. 2009b] developed an MDA Secure Engineering Process for DWs. So it opens new research areas regarding developing CMMI, ISO or PMI based MDA DW development methods.

5. FUTURE RESEARCH TRENDS

After browsing the current research direction in MDA and DW and highlighting the current open research problems, we summarize the future research trends as follow:

Traditional DWS

1) Creating new Performance UML profile that targets the conceptual representation of performance measures for DW. This profile will be used to describe all elements needed to make a DW system react in efficient way. Profile elements may include these capabilities:
   - Enable data partitioning with different partitioning types such as "hash function" or "Rang of values".
   - Enable Data indexing with different indexing types such as bitmap index, function index
   - Enable automatic creation of materialized views for fast data retrieval.
2) Creating an MDA CMMI, ISO, and PMI based data warehouse development framework that handles MDA technical issues side by side with managerial issues.
3) Creating new UML profiles to build MDA conceptual models to generate the data mining, what-if-analysis OLAP and ETL application for traditional DWs.
4) Extending UML to consider new stereotypes regarding object-oriented and object-relational databases for an automatic generation of the database schema into these kinds of databases.
5) Proposing a group of metrics as a means to describe good MD models based on more objective criteria.

Spatial DWs

6) Improving existing MDA approaches for the development of Geographical DWs by adding other applications metadata generation such as data mining and what-if-analysis.
7) Improving MDA approach for the development of Spatial DWs by adding other PSMs according to several platforms (e.g. object and object relational platforms).
8) Extending the spatial elements presented to some complex levels and measures such as temporal measures, time dimensions.
9) Developing secure GDW profile.
10) Applying MDA concepts to a spatial data mining application.
11) Using MDA for building GDW ETL programs.

Secure DWs

12) Creating a formal MDA transformation by using QVT or ATL between secure CIM and secure PIM.
13) Working on developing several secure PSMs, such as secure Multidimensional Online Analytical Processing (MOLAP) and secure Hybrid Online Analytical Processing (HOLAP).
14) Adapting the Model-to-Text approach in order to transform models into code for specific DBMS such as Oracle, SQL Server or MySQL.
15) Building a CASE tool developed in order to automatically implement secure DWs.

CWM and XMI

16) Creating CWM-based DW building tool which eases the process of metamodel transfer between different environments that support all CWM specifications for modeling metadata for relational, non-relational, multidimensional systems.

6. CONCLUSION

We have provided a comprehensive survey highlighting current progress of the exciting topic of Model Driven Architecture (MDA) and Data warehousing. To understand the field of MDA and DW, we surveyed research trends in MDA and DW using ACM, IEEE, Science Direct and Springer journals that are relevant to MDA and DW research. Firstly MDA research focused on automating the creation of Multidimensional model (Start schema) from Conceptual Models. Secondly, creation of new UML profiles to address new business domain such as security for DW; then using this profile to furnish their UML models. Thirdly, they used MDA to model the non-functional requirement of DW in early stages of SDLC side by side with functional requirement; e.g. Security—authorization, authentication. However, there is room to improve further the DW development with MDA concepts. This paper highlights new research directions within MDA and DW field, which could improve DW solution quality and performance and also minimize drawbacks and limitations. We discuss potential research areas such as using MDA concepts to model and automate the creation of DW performance parameters such as creating materialized view, bitmap index, data partitioning. Furthermore there is a space for applying MDA concepts to other stages of DW development such as ETL and OLAP end user Application stages. Other than that and to the best of our knowledge all the MDA-based DW development methodologies are technical oriented focusing on model-to-model generation; Managerial challenges for DW have not been articulated. As a long-term goal of research, we believe having a complete self-contained MDA-based methodology that handles technicalities and managerial issues in one box is a great contribution to the field.

REFERENCES

ANNEKE KLEPPE, JOS WARMER, WIM BAST. 2003. *MDA Explained the Model-Driven Architecture: Practice and Promise.* Boston: Addison-Wesley.

BEDIR TEKINERDOGAN, MEHMET AKSIT AND FRANCIS HENNINGER. 2007. Impact of Evolution of Concerns in the Model-Driven Architecture Design Approach. Electronic Notes in Theoretical Computer Science. 163, 2, 45-64.

BIMONTE, S., TCHOUNIKINE, A., MIQUEL, M. 2005. Towards a spatial multidimensional model. In the Proceedings of the 8th ACM international workshop on Data warehousing and OLAP(DOLAP 2005) (Bremen, Germany). November 04 - 05.

BRAN SELIC. 2007. A Systematic Approach to Domain-Specific Language Design Using UML. In the proceedings of the 10th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC'07) (Santorini Island, Greece). May 07- 09.

CAROL C. BURT, B.R. BRYANT, R.R. RAJE, A.M. OLSON, AND M.AUGUSTON. 2003. Model Driven Security: Unification of Authorization Models for Fine-Grain Access Control. In the Proceedings of the 7th International Enterprise Distributed Object Computing Conference (EDOC'03). Brisbane, Australia, September 16-19. 159-173.

CHUNG, L., NIXON, B., YU, E., MYLOPOULOS, J. 2000. Non-Functional Requirements in Software Engineering. Kluwer Academic Publishers. Boston Hardbound.

DANIAL T. CHANG. 2001. CWM Enablement Showcase. www.cwmforum.org. Accessed July 2009.

DAVID BASIN, JÜRGEN DOSER, AND T. LODDERSTEDT. 2006. Model Driven Security: From UML models to access control infrastructures. ACM Transactions on Software Engineering and Methodology (TOSEM). 15, 1, 39 - 91.

DENNIS M. AHERN, AARON CLOUSE , RICHARD TURNER. 2008. CMMI Distilled: A Practical Introduction to Integrated Process Improvement,Third Edition. Addison Wesley Professional.

EDUARDO FERNANDEZ-MEDINA, J. TRUJILLO, R. VILLARROEL, AND M.PIATTINI. 2006a. Access control and audit model for the multidimensional modeling of Data Warehouses. DSS. 42, 1270-1289.

EDUARDO FERNANDEZ-MEDINA ET AL. 2006b. Developing secure Data Warehouses with a UML extension. Information Systems. 32, 6, 826-856

ELZBIETA MALINOWSKI AND ESTEBAN ZIMÁNYI. 2004a. OLAP Hierarchies: A Conceptual

Perspective. Advanced Information Systems Engineering, LNCS 3084, 477–491

ELZBIETA MALINOWSKI, ZIM´ANYI, E. 2004b. Representing spatiality in a conceptual multidimensional model. In the Proceedings of the 12th annual ACM international workshop on Geographic information systems(GIS 2004) (New York, USA). November 12-13.

ELZBIETA MALINOWSKI AND ESTEBAN ZIMÁNYI. 2006. A conceptual solution for representing time in Data Warehouse dimensions. In the Proceedings of the 3rd Asia-Pacific conference on Conceptual modeling (Hobart, Australia). January 16-19. 45-54.

EMILIO SOLER,VILLARROEL R., TRUJILLO J., PIATTINI, FERNÁNDEZ-MEDINA. 2006. Representing security and audit rules for DW at the logical level by using the CWM. In the Proceedings of The First International Conference on Availability, Reliability and Security(ARES2006) (Vienna, Austria). April 20-22.

EMILIO SOLER, JUAN TRUJILLO, EDUARDO FERNÁNDEZ-MEDINA AND MARIO PIATTINI. 2007a.A Framework for the Development of Secure Data Warehouses based on MDA and QVT. In the Proceedings of the second International Conference on Availability, Reliability and Security (ARES'07) (Vienna , Austria). April 10-13.

EMILIO SOLER, VERONIKA STEFANOV, JOSE-NORBERTO MAZ´ONZ, JUAN TRUJILLO, EDUARDO FERN ANDEZ-MEDINA AND MARIO PIATTINI. 2008a. Towards Comprehensive Requirement Analysis for Data Warehouses: Considering Security Requirements. In the Proceedings of the Third International Conference on Availability, Reliability and Security, (ARES 2008) (BARCELONA, SPAIN). March 4-7.

EMILIO SOLER, J. TRUJILLO, E. FERNÁNDEZ-MEDINA, M. PIATTINI. 2008b. Building a secure star schema in Data Warehouses by an extension of the relational package from CWM. Computer Standards & Interfaces. 30, 6, 341–350.

IL-YEOL SONG, RITU KHARE, YUAN AN, SUAN LEE, SANG-PIL KIM, JINHO KIM AND YANG-SAE MOON. 2008. SAMSTAR: An Automatic Tool for Generating Star Schemas from an Entity-Relationship Diagram. LNCS. 5231, 522–523.

INMON,B. 2005. Building the Data Warehouse. Wiley Publishing Inc. Fourth Edition.

ISO. 2009. INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. http://www.iso.org. Access on June 2009.

JESUS PARDILLO, JOSE-NORBERTO MAZÓN, JUAN TRUJILLO. 2008. Bridging the semantic gap in OLAP models: platform-independent queries. In the Proceedings of the DOLAP'08 (Napa Valley, California, USA). October 30.

JOHN P, DAN CHANG, S. TOLBERT AND DAVID M. 2003. Common Warehouse Metamodel Developer's Guide. Wiley publishing. Indiana.

JOSE-NORBERTO MAZÓN, JUAN TRUJILLO, MANUEL SERRANO, MARIO PIATTINI. 2005. Applying MDA to the Development of Data Warehouses. In the Proceedings of the DOLAP'05 (Bremen, Germany). November 04–05.

JOSE-NORBERTO MAZON, J. TRUJILLO, ORGER LECHTENB. 2007a. Reconciling requirement driven Data Warehouses with data sources via multidimensional normal forms. Data & Knowledge Engineering. 63, 3, 725-751.

JOSE-NORBERTO MAZON, JESUS PARDILLO, JUAN TRUJILLO. 2007b. A model-driven goal-oriented engineering approach for Data Warehouses.. In the Proceedings of the International Workshop on Requirements, Intentions and Goals in Conceptual Modeling (ER Workshops 2007) (Auckland, New Zealand), November 5–9.

JOSE-NORBERTO MAZON, JUAN TRUJILLO. 2007c. A Model Driven Modernization Approach for Automatically Deriving Multidimensional Models in Data Warehouses.. In the Proceedings of the 6th International Conference on Conceptual Modeling (ER 2007) (Auckland, New Zealand). November 5–9.

JOSE-NORBERTO MAZON, J. TRUJILLO. 2008. An MDA approach for the development of Data Warehouses. Decision Support Systems. 45, 1, 41–58.

JUAN TRUJILLO, EMILIO SOLER, EDUARDO FERNÁNDEZ-MEDINA, MARIO PIATTINI. 2009a. A UML 2.0 profile to define security requirements for Data Warehouses. Computer Standards & Interfaces. 31, 5, 969-983.

JUAN TRUJILLO, EMILIO SOLER, EDUARDO FERNANDEZ-MEDINA, MARIO PIATTINI. 2009b. An engineering process for developing Secure Data Warehouses. Information and Software Technology. 51, 1033–1051.

KIMBALL, R. 2002. The Data Warehouse Toolkit. Second Edition. Wiley Publishing.

KRZYSZTOF J. CIOS, ROMAN W. SWINIARSKI, WITOLD PEDRYCZ AND LUKASZ A. KURGAN. 2007. Databases, Data Warehouses, and OLAP. Data Mining A Knowledge Discovery Approach. Springer US. 95-131.

KUMPON FARPINYO AND TWITTIE SENIVONGSE. 2003. Designing and creating relational schemas with a CWM-based tool. In the Proceedings of the 1st international symposium on Information and communication technologies (ISICT '03) (Dublin, Ireland). September 24 – 26.

LEOPOLDO ZEPEDA, MATILDE CELMA. 2005. Specifying Metamodel transformations for Data Warehouse design. In the Proceedings of the ACM Southeast Conference (Kennesaw, GA, USA).

LILIA MUNOZ, JOSE-NORBERTO MAZON, JESUS PARDILLO AND JUAN TRUJILLO. 2008. Modeling ETL Processes of Data Warehouses with UML Activity Diagrams. OTM 2008 Workshops, LNCS, 5333, 44–53.

LYNN LANGIT. 2007. OLAP Modeling. Foundations of SQL Server 2005 Business Intelligence. Apress. 25-49.

MARK RITTMAN. 2009. Using Oracle Business Intelligence Discoverer with the OLAP Option. http://www.oracle.com/ technology/ pub/ articles/rittman_olap.html. Accessed on Aug 24 2009.

MIKKO KONTIO, 2005. Architectural manifesto: The MDA adoption manual. http://www.ibm.com/developerworks/ wireless/library/wi-arch17.html. Accesed on Aug 22, 2009.

MING CHUAN HUNG, MAN-LIN HUANG, DON-LIN YANG, NIEN-LIN HSUEH. 2007. Efficient approaches for Materialized Views selection in a Data Warehouse. Information Sciences. 177, 1333–1348.

MIT. 2003. META INTEGRATION TECHNOLOGY INC. Meta Integration Model Bridge. http://www.metaintegration.net/ Products/ MIMB/. Accessed on July 2009.

NAVEEN PRAKASH, YOGESH SINGH, AND ANJANA GOSAIN. 2004. Informational scenarios for Data Warehouse requirements elicitation. In the Proceedings of the ACM 23rd International Conference on Conceptual Modeling (Shanghai, China). November 8- 12.

OCTAVIO GLORIO, JUAN TRUJILLO. 2008. An MDA Approach for the Development of Spatial Data Warehouses. In the Proceedings of the ACM 10th international conference on Data Warehousing and Knowledge Dis (Turin, Italy). LNCS. 5182, 23-32.

OCTAVIO GLORIO, JUAN TRUJILLO. 2009. Designing Data Warehouses for Geographic OLAP Querying by using MDA. LNCS Computational Science and Its Applications. 5592, 505-519.

OGC. 2009. OPEN GEOSPATIAL CONSORTIUM. Simple Features for SQL. http://www.opengeospatial.org. Accessed on July 2009.

OMG. 2003a. OBJECT MANAGEMENT GROUP. Common Warehouse Metamodel (CWM) Specification,V1.1, Volum 1.

OMG. 2003b. OBJECT MANAGEMENT GROUP. MDA Guide Version 1.0.1.

OMG. 2003c. OBJECT MANAGEMENT GROUP. Response to the UML 2.0 OCL RFP, OMG document ad/03-01-07.

OMG. 2005a. OBJECT MANAGEMENT GROUP. Software & Systems Process Engineering MetaModel Specification, version. 1.1, http://www.omg.org/docs/formal/05-01-06.pdf.

OMG. 2005b. OBJECT MANAGEMENT GROUP. UML Superstructure Specification, v2.0. http://www.omg.org/cgi-bin/doc?formal/05-07-04.

OMG. 2007. OBJECT MANAGEMENT GROUP. XML Metadata Interchange (XMI).

OMG. 2008. OBJECT MANAGEMENT GROUP. Meta Object Facility (MOF) 2.0 Query/ View/ Transformation Specification Version 1.0. OMG Document Number: formal/2008-04-03.

OMG. 2009. OBJECT MANAGEMENT GROUP. Unified Modeling Language(UML), Infrastructure Version 2.2. http://www.omg.org/spec/UML/2.2/Infrastructure/PDF.

PAIM FÁBIO RILSTON SILVA AND JAELSON F. B. CASTRO. 2002. Enhancing Data Warehouse Design with the NFR Framework. In the Proceedings of the fifth Workshop on Requirements Engineering (WER 2002)(Valencia, Spain). November 11-12.

PAOLO GIORGINI, STEFANO RIZZI, MADDALENA GARZETTI. 2005. Goal-oriented requirement analysis for Data Warehouse design. In the Proceedings of the DOLAP'05 (Bremen, Germany). November 04–05.

PETER BELLSTR◌M. 2005. Using Enterprise Modeling for Identification and Resolution of Homonym Conflicts in View Integration. Information Systems Development. Springer US. 265-276.

POSTGIS. 2009. Refractions Research : PostGIS, version 1.4. http://postgis.refractions.net/. Accessed on July 2009.

POSTGRESQL. 2009. Global Development Group. PostgreSQL, version 8.4. http://www.postgresql.org. Accesed July 2009.

RACHID ANANE. 2001. Data Mining and Serial Documents. Computers and the Humanities. 35, 299–314.

RODOLFO VILLARROEL, EDUARDO FERNÁNDEZ-MEDINA, MARIO PIATTINI, JUAN TRUJILLO. 2006. A UML 2.0/OCL Extension for Designing Secure Data Warehouses. Journal of Research and Practice in Information Technology. 38, 1.

ROMERO, O., ABELL´O, A. 2007. On the need of a reference algebra for OLAP. In the Proceedings of the 9th International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2007) (Regensburg, Germany). September 3-7.

SERGIO MORA, AND JUAN TRUJILLO. 2002a. Extending UML for Multidimensional Modeling. In the Proceedings of the 5th International Conference on the Unified Modeling Language (UML 2002)(Dresden, Germany). September 30– October 4.

SERGIO MORA , JUAN TRUJILLO , YEOL SONG. 2002b. Multidimensional modeling with UML package diagrams.In the Proceedings of the 21st Intl. Conference on Conceptual Modeling (Tampere, Finland). October 7-11.

SERGIO MORA, JUAN TRUJILLO, YEOL SONG. 2006. A UML profile for multidimensional modeling in Data Warehouses. Data & Knowledge Engineering. 59, 725–769.

SIVANANDAM.S.N., G.R. KARPAGAM. 2004. A Novel approach for Implementing Security services. Academic Open Internet Journal, 13.

TORBEN BACH PEDERSEN. 2009. Warehousing The World: A Vision for Data Warehouse Research. New Trends in Data Warehousing and Data Analysis. Springer US, 3, 1-17.

WEN-YANG LIN1 AND I-CHUNG KUO. 2004. A Genetic Selection Algorithm for OLAP Data Cubes. Knowledge and Information Systems. 6, 1,83-102

WINTER R. , STRAUCH B. 2003. A method for demand-driven information requirements analysis in data warehousing projects. In the Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS-36 2003) (Big Island, HI, USA.). January 6-9.

YU, ERIC. Towards Modelling and Reasoning Support for Early-Phase Requirements Engineering. 1997. In the Proceedings of the 3rd Int. Symp. on Requirements Engineering (Annapolis, USA.). January 5-8.

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 3, March 2016

A Framework for Building Ontology in Education Domain for Knowledge Representation

Monica Sankat [#1], R S thakur [*2], Shailesh Jaloree [#3]

[#] *Department of Applied Mathematics and Computer Application, SATI*
Vidisha, India

[*] *Department of Computer Application, MANIT*
Bhopal, India

***Abstract*—In this paper we have proposed a method of creating domain ontology using protégé tool. Existing ontology does not take the semantic into context while displaying the information about different modules. This paper proposed a methodology for the derivation and implementation of ontology in education domain using protégé 4.3.0 tool.**

I. Introduction

There is large amount of data available on net, which is dispersed, superfluous and inaccurate by nature, makes use of the information difficult. This problem is often referred to as Information overload. Existing technologies lack ability to perform significant analysis and filtering of data, there by presenting results that only human can process and not machine.

The purpose of semantic web idea was to provide meaningful web that can be processed by machines and humans equally [1]. The web can review the intent of user and provide results that fulfil the information requirement. Since, there is a prospective to create diverse ontologies on a same domain as no common criteria exist for building ontologies. This paper presents a methodology for the derivation and implementation of ontology in education domain. The key concepts of the domain with its data properties have been discussed. Model is implemented in using protégé 4.3.0. This paper covers the major aspects of Education domain including super class and subclass hierarchy, creating a subclass instances for class diagram, properties and their relations etc.

II. RELATED WORK

WebODE [2] is an advanced ontological engineering workbench that provides varied ontology related services, and gives assistance to most of the activities involved in the development of ontology.

Ontology pruning is to build a domain ontology based on different heterogeneous sources. It has the following steps. First, for the domain-specific ontology, core ontology is used as a top level organization. Second, a dictionary is used to acquire domain concepts. Third, concepts that were not domain specific are removed by domain specific corpora of texts [3].

Protégé [4] is probably the most popular ontology development tool. Protégé is a free, Java-based open source ontology editor. Protégé offers two approaches for the modelling of ontologies: a traditional frame-based approach (via Protégé-Frame) and a modelling approach using OWL (via Protégé-OWL). Protégé ontologies can be stored in a variety of different formats, including RDF/RDFS, OWL and XML Schema formats.

Arabshian [5] in his paper propose LexOnt; a semi-automatic ontology creation tool for a high-level ontology. LexOnt explores Web directory as corpus, although it can evolve to use other corpora as well.

LexOnt [6] is developed as a Protege plug-in .The GUI design and implementation of LexOnt. LexOnt is built specifically for those who are not experts within a domain, but for users who want to recognize the domain on a high-level and create an ontology that describes it.

Boyce [7] presented a method for domain experts to develop ontologies for use in the delivery of courseware content. They focused in particular on relationship types that allow us to represent rich domains sufficiently.

Fortuna [8] proposed a semi-automatic and data-driven ontology editor called OntoGen, focusing on editing of topic ontologies .The system combines text data mining techniques with an efficient user interface to decrease the time spent and complexity.

Fortuna [9] presents a new version of OntoGen system. The system integrates machine learning and text data mining algorithms into an efficient user interface making ease of use for users who are not ontology engineers.

Mei-ying Jia et al. [10] has proposed automated ontology construction method. The method is not pure auto-mated. It uses existing thesaurus and database of Military Intelligence. The thesaurus provides classes information for the ontology and the database provides the instances. Here, only three types of relationships are used between concepts of constructed ontology.

Bhowmick [11] present a framework for manual ontology engineering in education domain for managing learning content of the syllabus related requirements of school students. In this paper, a multilingual framework for management of knowledge structures of such domains.

To reduce the effort of manual ontology building, Choudhary propose a methodology for building ontology in semi-automatic manner. In his paper algorithms are developed for automatic discovery of concepts from Web for building domain ontology. Relationships among the concepts are assigned in semi-automated manner [12].

Navigli [13] in his paper presented a methodology for automatic ontology enrichment and document explanation with concepts and relations of an existing ontology. They defined Natural language definitions from available taxonomies in a given domain are processed. These regular expressions are useful to identify general-purpose and domain-specific relations.

III. The Domain Ontology Problem

The nature of ontology changes domain to domain. Steps will be taken up into concern for building ontology for Education domain, same steps likely would not consider for structure ontology for some other domain like education, finance, health care etc because the nature of domain in some cases top to bottom or vice versa. The main drawbacks in existing work in this area are:-

- There are not integrated methods and tools that combine different techniques and diverse knowledge sources with existing ontologies to accelerate the development process and these methods are not generalized to other domains [11].
- They only provide some specialized relationships among the concepts. Again these relationships are not adequate to describe knowledge constitute of education domain [15].
- Doesn't provide Easy interface for domain experts having little technological expertise.

IV. Motivation

The specific features of this domain are:-

- Every concept refers to a semantically distinct entity. The concepts in a domain are related to each other through different relationships.
- Different types of relationships may exist and the same concept can be represented by different words.
- The phenomenon of synonymy is very common. So the same concept may be referred to by several terms. For example, the terms DM and Data Model refers to the same concept.

Following Fig.1. Shows the Domain Ontology taxonomy:

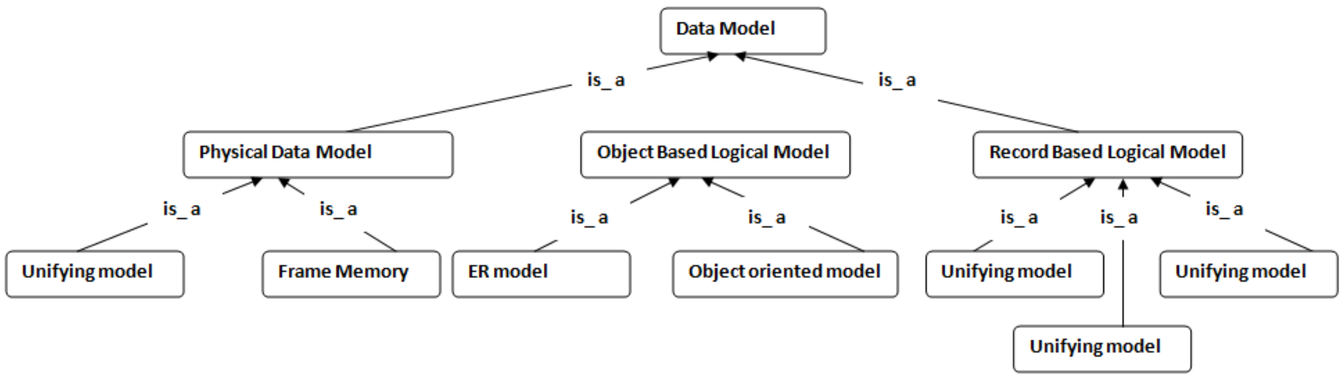


Fig .1: "Data Model" Taxonomy

V. Proposed work

The main initiative in this paper is to research and characterize appropriate approach to ontology development. Analysing the specific features of the domain, it is identified that requirements for representing the domain knowledge are as follows:-

- Representing the Educational domain which can serve to potential students in making the choice of their desirable Concepts.
- Creating Meta data about Educational systems.
- Reducing the Redundancy occurring due to the synonymous ambiguity between the terms to find information at the concept level is very significant
- Different types of relationships may be used in many ways in systems that make use of the domain knowledge.

VI. Ontology Building Method

Building domain-specific ontologies is an expensive construction task. This approach is to develop domain ontology for educational data. Ontology is built in this step by linking concepts and relations extracted. The proposed work will focus on relationship types that allow us to model rich domains effectively. The method used to build Ontology is shown in Fig.1.

A. Acquirement of Ontology using Text mining

Text mining, also known as Intelligent Text Analysis, Text Data mining is a process of extracting interesting and non-trivial information and knowledge from unstructured text [14]. Knowledge may be discovered from many sources of information but there are many unstructured texts remain as largest source of knowledge. The problem of Text Data Mining is to acquire implicit and explicit concepts and semantic relations between concepts using Natural Language Processing (NLP) techniques.

B. Filtering of Domain Ontology

The next step is to convert the collected text documents (in unstructured form) to a structured .Parsing is the first step in converting unstructured text to the structured format for ease of analysis. Typically, this process involves tokenization, normalization of tokens (lemmatization or stemming), Part – of-speech (POS) tagging and so on [10].

C. Extraction of concepts

In this step, concepts i.e. domain oriented terms are extracted. For example, Object, Attributes, Entities and Data models. Occurrences of Term and their Word Count is also calculated i.e. Occurrence of Term “Data models” in following sentence, “Object based data models has concepts such as entities, attributes, and relationships” is 1 and its Word Count is 2.

D. Identification of Relationship among the concept

For the topic of data Model in the education domain in which most relationships between the concepts is shown by ‘is-a’ relationship, there are some relationships between concepts that which are not generalization or specialization relationship types and hence if the ‘is-a’ relationship was used, the relationships would be misinterpreted. Hence in “Data Model” Ontology, a number of other relationship types were created and defined such as Has_Part, Has_Subtype.

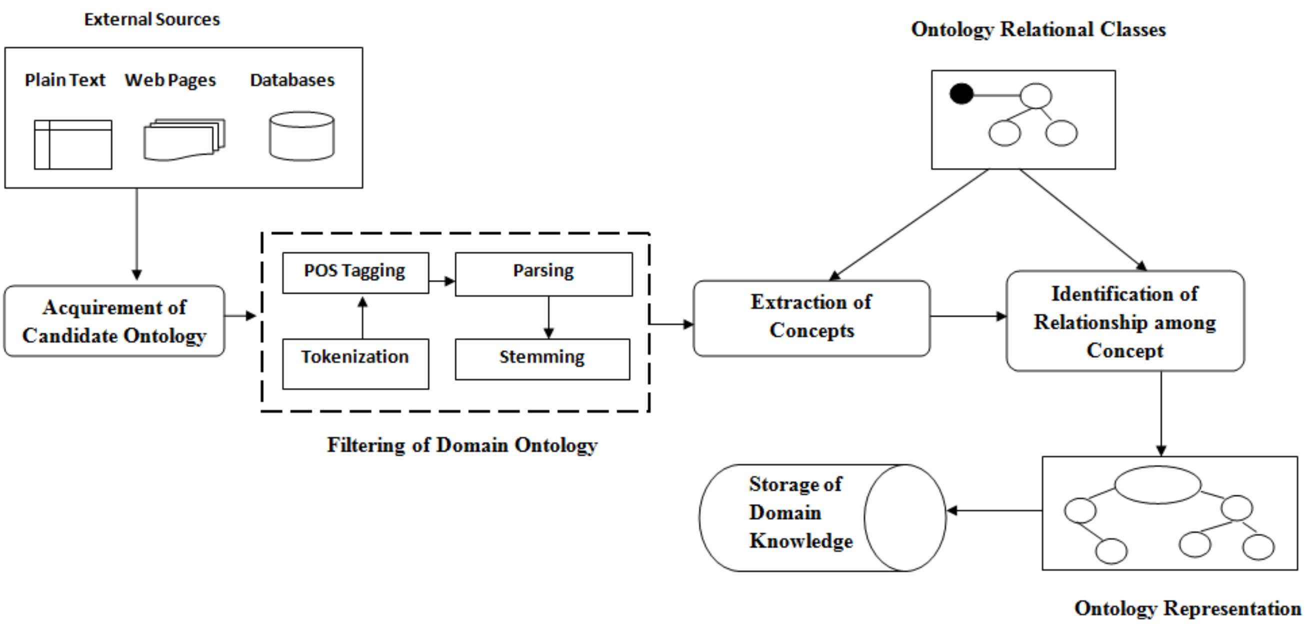


Fig .2: Ontology Building Framework

VII. IMPLEMENTING THE EDUCATIONAL ONTOLOGY WITH PROTÉGÉ 4.3.0

In order to implement the ontology, we chose Protégé 4.3.0 because of the fact that it is extensible and provides a user friendly environment. In the following section ontology Classes, their Object properties and their Disjoint Classes are shown.

A. Classes and class hierarchy

The first step was to give the “Data model” related classes or concepts. Further the concepts are mainly divided into Physical, Object based and Record based, as shown in Fig. 3.

B. Disjoint Classes

If classes cannot have any common instances they are called Disjoint Classes. Disjoint classes for “Data Model” ontology are shown in Fig. 4.

C. Object properties of ontology

Object properties for representing the relationships which we want to add among classes are shown in Fig. 5

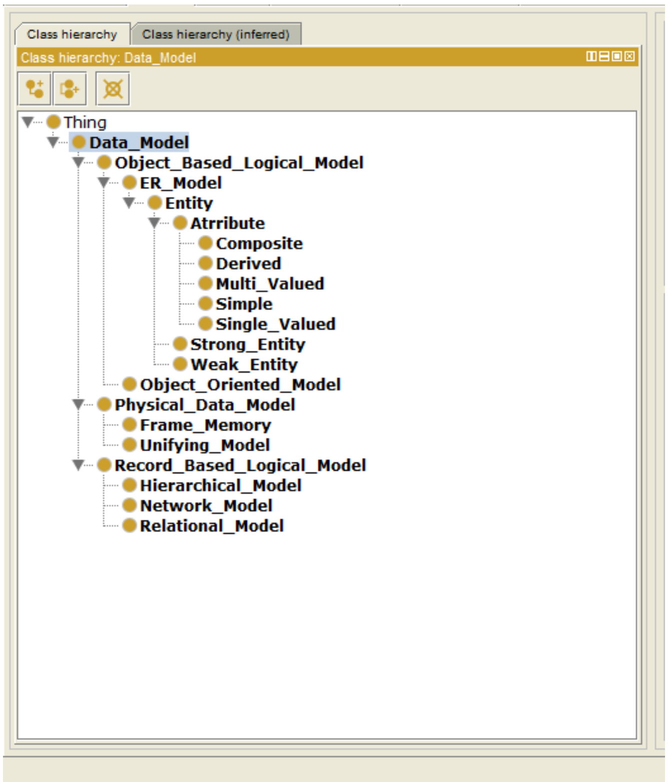


Fig.3: Class Hierarchy Representation of "Data Model"

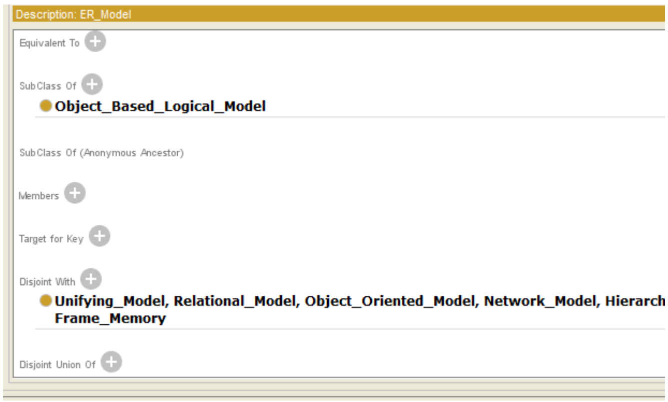


Fig.4: Disjoint Classes in Protégé 4.3.0

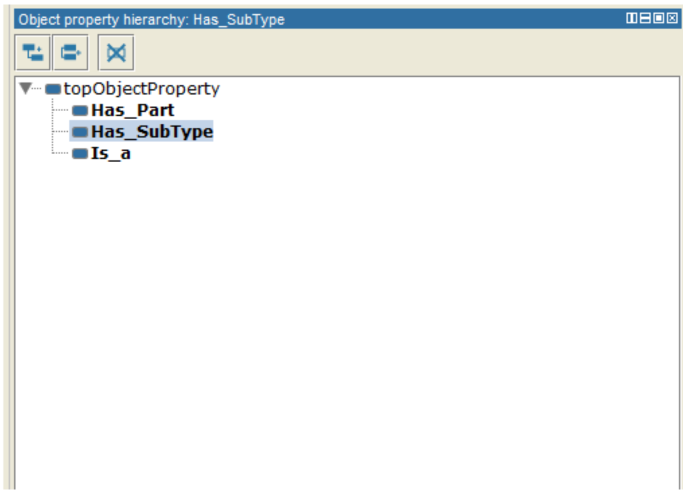


Fig .5: Object Properties in Protégé 4.3.0

VIII. VISUALIZATION OF ONTOLOGY

The final process that generates an ontology as the knowledge representation is shown in Fig 6 .In this ontology we add important classes and subclasses of "Data model" Ontology as shown in Fig. 6.

IX. CONCLUSION

This paper details the steps that transform taxonomy into a domain concept and explains how this structure is transformed into more formal domain ontology. We would like to realize the generation of more complex concepts that exploit the existing ontology concept as well as available ontologies to fulfill Educational objective.

REFERENCES

[1] "W3C Semantic Web Activity". World Wide Web Consortium (W3C). http://www.w3.org/2001/sw/. Retrieved Jan 15, 2012.

[2] Vega, J. C. A. (2000). WebODE 1.0: User's Manual. Laboratory of Artificial Intelligence,Technical University of Madrid

[3] Kietz ,JU., Maedche ,A., Volz, R. (2001), "A Method for Semi-Automatic Ontology Acquisition from a Corporate Intranet," In: N. Aussenac-Gilles,B. Biebow, S. Szulman(eds) EKAW'00 Workshop on Ontologies andTexts. Juan-Les-Pins, France. CEUR Workshop Proceedings 2000,51:4.1-4.14. Amsterdam, The Netherlands Available: http://CEURWS.org/Vol-51 ,2000

[4] Gennari, J. H., Musen, M. A., Fergerson, R. W., Grosso, W. E., Crubzy, M., Eriksson, H.,Noy, N. F., Tu, S.W. (2002). The evolution of Protege: an Environment for Knowledgebased Systems Development. International Journal of Human Computer Studies, 58, 89–123.

[5] Arabshian, K., Danielsen, P. & Afroz, S.(2012). Lexont: Semiautomatic ontology creation tool for programmable web. In AAAI 2012 Spring Symposium on Intelligent Web Services Meet Social Computing, Palo Alto, CA, March 2012.

[6] Danielsen, Peter J. & Arabshian, Knarig (2013). User Interface Design in Semi-Automated Ontology Construction,IEEE 20th International Conference on Web Services,2013.

[7] Boyce, S., & Pahl, C. (2007). Developing Domain Ontologies for Course Content. Educational Technology & Society, 10 (3),275-288.

[8] Fortuna Blaz, Grobelnik Marko & Mladenic Dunja(2006): Semi-automatic Data-driven Ontology Construction System. In: Proceedings of the 9th International multi-conference Information Society IS-2006, Ljubljana, Slovenia (2006).

[9] Fortuna Blaz, Grobelnik Marko & Mladenic Dunja(2007), OntoGen: Semi-automatic Ontology Editor: Human Interface, Part II, HCII 2007, LNCS 4558, pp. 309–318, 2007.

[10] Jia, M., Yang, B., Zheng D., Sun, W., Liu, Li., Yang, Jing.,(2009) "Automatic Ontology Construction Approaches and Its Application on Military Intelligence", Asia-Pacific Conference on InformationProcessing (APCIP), vol. 2, Pp. 348 – 351, 2009.

[11] Bhowmick P.K., Roy D., Sarkar S. & Basu A.(2010), A Framework For Manual Ontology Engineering For Management Of Learning Material Repository,2010.

[12] Choudhary Jaytrilok. & Roy Devshri.(2012) ,An Approach to Build Ontology in Semi-Automated way, Journal Of Information And Communication Technologies, Volume 2, Issue 5, May 2012.

[13] Navigli, R., Gangemi, A., Velardi, P.(2003). Ontology learning and its application,2003

[14] Haralampos Karanikas and Babis Theodoulidis Manchester, (2001), "Knowledge Discovery in Text and Text Mining Software", Centre for Research in Information Management, UK.

[15] Zhou, W., Liu, Z., Zhao, Y., Xu, L., Chen, G., Qiang, W., Huang, M. and Qiang, Y. 2006. A Semi-automatic Ontology Learning Based on Word Net and Event-based Natural Language Processing. In: International Conference on Information and Automation

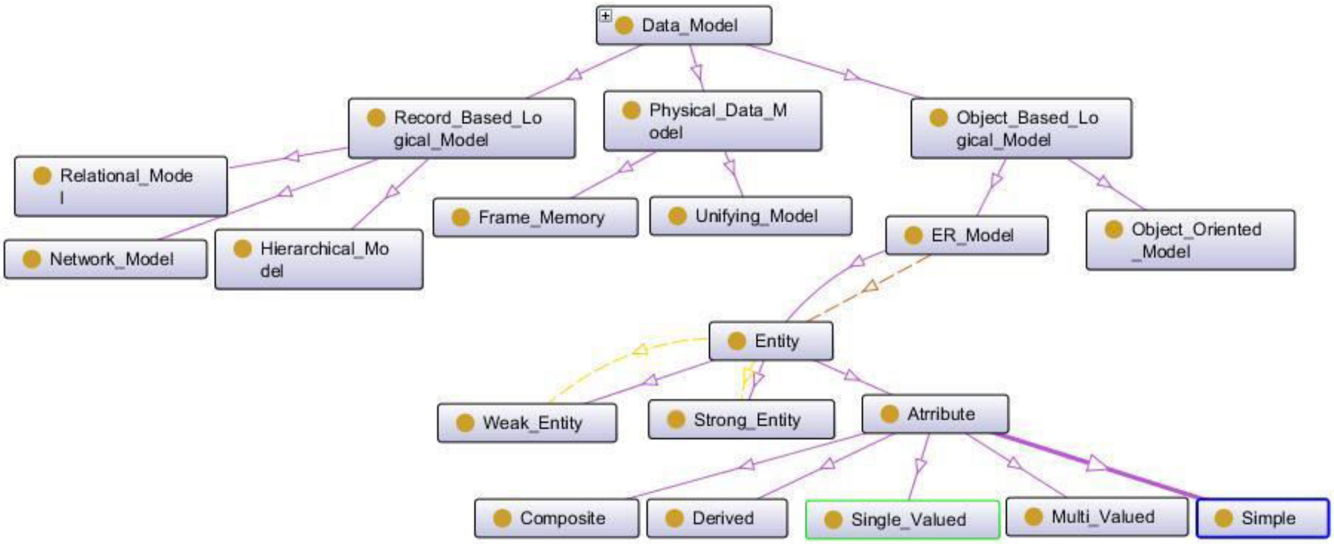


Fig.6: Visualization of Ontology

Mobility Aware Multihop Clustering based Safety Message Dissemination in Vehicular Ad-hoc Network

Nishu Gupta, Arun Prakash, Rajeev Tripathi

Abstract- **A major challenge in Vehicular Ad-hoc Network (VANET) is to ensure real-time and reliable dissemination of safety messages among vehicles within a highly mobile environment. Due to the inherent characteristics of VANET such as high speed, unstable communication link, geographically constrained topology and varying channel capacity, information transfer becomes challenging. In the multihop scenario, building and maintaining a route under such stringent conditions becomes even more challenging. The effectiveness of traffic safety applications using VANET depends on how efficiently the Medium Access Control (MAC) protocol has been designed. The main challenge while designing such a MAC protocol is to achieve reliable delivery of messages within the time limit under highly unpredictable vehicular density. In this paper, Mobility aware Multihop Clustering based Safety message dissemination MAC Protocol (MMCS-MAC) is proposed in order to accomplish high reliability, low communication overhead and real time delivery of safety messages. The proposed MMCS-MAC is capable of establishing a multihop sequence through clustering approach using Time Division Multiple Access mechanism. The protocol is designed for highway scenario that allows better channel utilization, improves network performance and assures fairness among all the vehicles. Simulation results are presented to verify the effectiveness of the proposed scheme and comparisons are made with the existing IEEE 802.11p standard and other existing MAC protocols. The evaluations are performed in terms of multiple metrics and the results demonstrate the superiority of the MMCS-MAC protocol as compared to other existing protocols related to the proposed work.**

Nishu Gupta, Arun Prakash and Rajeev Tripathi are with the Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad-211004, India

Index Terms **-- Clustering, Multihop, Safety, TDMA, V2V, VANET**

I. INTRODUCTION

Vehicular Ad-hoc Network (VANET) provides vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication in order to support safety, traffic-management and non-safety applications. The messages exchanged by safety applications require predictable or low delay and high reliability. Even a slight delay in delivery of messages may significantly affect the performance of safety applications. In addition, safety messages have a time bound deadline and the message should reach the destination within this time limit. In particular, the effectiveness of active safety applications depends on the ability to disseminate messages as quickly as possible with high reliability, fairness and scalable utilization of network resources [1]. In contrast to the contention-based protocol such as IEEE 802.11, clustering-based Time Division Multiple Access (TDMA) scheme attracts more attention for VANET to improve the traffic safety applications as the number of nodes is increased to a large number [2]. The approved amendments to the IEEE 802.11 standard standardized as IEEE 802.11p or Wireless Access in Vehicular Environments (WAVE) has inherent shortcomings of not being able to provide reliable broadcast services. With the random channel access, it suffers from unbounded latency and broadcast storm [3]. Consequently, it experiences huge amount of packet loss, collisions and access delays. These challenging issues are intermittently associated with contention-based Medium Access Control (MAC) protocols. Another challenging task in the implementation of VANET is to design a Quality of Service (QoS) aware protocol. Such protocol would aim to alleviate delay while guaranteeing QoS constraints with respect to the Packet Delivery Ratio (PDR), throughput and reliable

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

message delivery. More than that, it would make efficient use of the network bandwidth. In order to overcome their challenging characteristics, most VANET routing algorithms use geographic based routing [4] and opportunistic carry-and-forward based routing techniques [4, 5]. These techniques leverage local or global knowledge of traffic statistics to implement multihop forwarding strategies in order to minimize communication overhead while adhering to delay constraints imposed by the application.

Many QoS parameters can be improved by using TDMA scheme with no central control so as to provide fair and reliable data dissemination in V2V communication [6]. Likewise, a vehicular scenario can be organized hierarchically using a clustering protocol. By clustering, the vehicles are partitioned into groups of minimum relative mobility to reduce the amount of routing information [7]. The most important criterion for any clustering method in VANET is to form stable clusters with minimum overheads. To realize this aim, nodes in the VANET are divided into different clusters based on their position, direction of movement, lanes and speed. In addition, the reliability of the safety messages is increased by assigning time slots to different nodes.

In this work, we focus on broadcasting of safety messages in the V2V scenario. Such messages demand high probability of successful delivery (PSD) and low latency, particularly in scenarios where there is no infrastructure support to coordinate communication. We propose Mobility aware Multihop Clustering based Safety message dissemination MAC (MMCS-MAC) protocol to increase reliability in VANET while delivering event-driven safety messages in multihop scenario over highway environment. Whereas many other schemes assume only a certain percentage of nodes to transmit safety message at any given point of time, the proposed scheme assures channel access to all the nodes, allowing them to transmit safety messages, no matter how severe application it demands. The novelty of the proposed algorithm lies in its dynamic adaptivity to mobility of the nodes and clustering based multihop forwarding strategies to achieve a good trade-off between delay and communication cost. This is in stark contrast with the previously proposed DMMAC [8], which aim at increasing the system's reliability, reducing the time delay for vehicular safety applications, and efficiently clustering nodes in highly dynamic and dense network in a distributed manner. An additional difference from the existing works is that no cluster head (CH) is required for allocating time slots to the nodes. This reduces an additional overhead and leverage in achieving high fairness.

We remark here the distinctive characteristics of our approach: (i) we apply multihop message routing scheme, up to four hops, so as to increase the message broadcast range in real-time event-driven applications (ii) we adopt mobility based clustering of nodes to increase the PDR and throughput of the safety messages (iii) we implement TDMA scheme to ensure reliability and fairness in the application (iv) we carry out division of the entire DSRC band into frames, and consider uniformly distributed vehicular density in order to achieve maximum channel utilization and (v) we do not require any changes to the existing IEEE WAVE stack.

Finally, we carry out extensive NS-2 simulations to evaluate the performance of MMCS-MAC with respect to different parameters. Results show that the proposed protocol outperforms several state-of-the-art protocols by achieving close to 100% reliability and faster dissemination, while the transmission overhead is much smaller.

The rest of the paper is organized as follows. Section 2 presents the related works, and in Section 3 we give the problem statement. Followed by that is the system model of the proposed protocol in Section 4. Section 5 evaluates and compares the performance of the proposed protocol with other related MAC protocols. Section 6 concludes the paper and discusses further direction of research.

II. RELATED WORKS

The main idea of cluster based routing scheme is to dynamically organize all mobile nodes into groups called clusters [9]. Support of QoS requirements in wireless ad-hoc network for distributed and real-time multimedia communication encounters a number of challenges, as specified in [10]. The authors in [11] design a cluster based aggregation–dissemination beaconing process that uses an optimized topology to provide nodes with a local proximity map of their vicinity. This would allow reliable inter-cluster bandwidth reuse during the aggregation phase. The topology is designed to minimize the inter-cluster interference by producing clusters that are separated by the maximal possible inter-cluster gaps. However, this optimization result proves to be less efficient for inter-cluster communication. Moreover, the probability of successful message reception decreases when node density increases in the intra-cluster communication. Evidently, this scheme would succumb to failure under high node density.

Several protocols have been proposed in VANET using TDMA to reduce interference and provide fairness between nodes. In [12], authors introduce a method for TDMA slot reservation based on clustering of vehicles, known as TDMA cluster based MAC (TC-MAC) for intra-cluster communications in VANET. TC-MAC integrates TDMA slot allocation with centralized cluster management technique. In this protocol, nodes are assigned time slots for collision free transmission. The work captivates on allowing vehicles to send and receive non-safety messages without any impact on the reliability of sending and receiving safety messages, even if the traffic density is high. In [13] the authors proposed a distributed mobility based clustering algorithm to increase cluster stability, where stability is realized by the time duration of the cluster members (CMs) and the CHs. These protocols generally use V2V communications for formation of clusters and for electing CHs.

In [14] the authors propose and evaluate a contention-free TDMA-based MAC approach that uses a predetermined multihop awareness range to distribute MAC slot allocation information to neighboring nodes. Nodes use the information from surrounding slots to select unused slots, thereby avoiding collisions. The multihop strategy is employed to overcome the hidden terminal problem. The results show that the optimal performance is achieved with two hops.

A multihop clustering scheme for VANET is proposed in [15]. To construct multihop clusters, a new mobility metric is introduced to represent relative mobility between nodes in multihop distance. The scheme highlights that multihop clusters can extend the coverage range of clusters and gain more advantages compared to single hop clusters. The work in [16] presents a clustering based MAC protocol designed to reduce interferences in VANET. The scheme is intended for safety applications in highway environments, employs dynamic multihop clustering and improves network performance. This approach also does not require cluster-head selection, similar to the algorithm proposed in this paper. A cluster based MAC (D-CBM) protocol is designed in [17] to ensure timely and reliable data delivery of messages. D-CBM employs distributed technique for clustering in VANET where V2V and V2I communication are considered. It is based on collision free TDMA in order to achieve high stability, low communication overheads and real time delivery of safety messages. In this protocol, the road side unit (RSU) assigns time slots to CHs and CH assigns time slots to CMs and gateway node. The RSU functions as a central coordinator to collect and distribute the messages. As the time slots can be assigned centrally, less number of collisions are expected which consequently increases the reliability. Other related works [18-23] have investigated the performance of safety-related applications based on metrics such as probability of successful delivery, end-to-end delay, forwarding node ratio, reachability, transmission overhead, transmission and receiver throughput, slot utilization rate etc.

III. PROBLEM STATEMENT

A. SYSTEM MODEL AND ASSUMPTIONS

In the system model, we focus on multihop transmissions where the clustered mobile nodes communicate via a single channel in purely ad-hoc mode. Each node within a cluster has a unique ID, based on its MAC address. Each node operates in the ad-hoc mode and broadcasts its packets according to a routing protocol. An example of V2V communication in multihop scenario has been depicted in Figure 1. It has been shown in [24] that Carrier Sense Multiple Access with collision avoidance (CSMA/CA) based MAC is not efficient enough to handle the high message frequency of the smart driving application. However, a TDMA-based MAC can be a probable solution to this issue.

We consider a 4-lane vehicular scenario and assume that each node is equipped with an IEEE 802.11p standard compliant radio device with GPS installed, which gives the location information. Each node shares information about its current position, speed, lane, and direction with only its one hop neighbors. To this aim, we impose that each safety message generated by a node within a cluster must be successfully delivered to all other clusters that are up to four hop counts (h_c), and in the direction of message travel.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

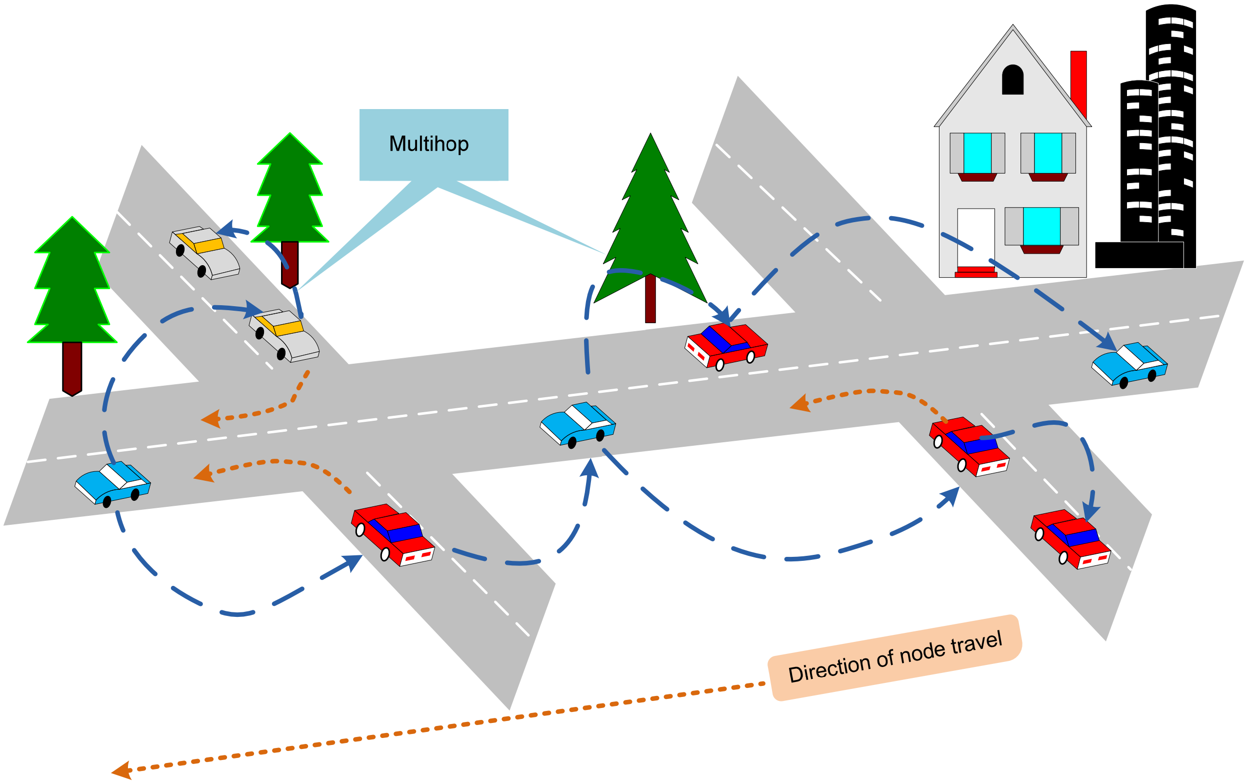


Figure 1. V2V communication in multihop scenario

B. OBJECTIVES

The MMCS-MAC protocol aims to achieve reliable dissemination and broadcast event-driven high priority safety messages by utilizing the entire DSRC band. It further aims to minimize the average delivery delay in the network by reducing interference in a long stream of vehicles. This would ensure real-time delivery of sensitive messages. In order to achieve our goal, we propose a TDMA based scheme designed for fast multihop channel access and a clustering mechanism that performs topology control and reduces interference while keeping the network connected. Data transmission of real-time safety messages is facilitated over IEEE 802.11 MAC-based channels in the allocated time slots.

IV. OVERVIEW OF MMCS-MAC PROTOCOL

MMCS-MAC protocol is divided into three different phases. The first phase is the cluster formation phase, where nodes are partitioned into different clusters according to their speed. The second phase constitutes the TDMA slot assignment. The aim of employing TDMA scheme on contention-based topology is to ensure reliable and fair transmission of safety messages. Realizing that safety related applications in vehicular communication urge for high reliability and low delay bound requirements, providing time to each node to transmit safety message without disturbing other nodes is crucial. TDMA stands out as the concept that can easily be used to allocate unique time slots to every cluster within the network. In the third phase, role of multihop forwarding in safety message dissemination comes into effect. Multihop forwarding refers to an aggressive message routing scheme where messages are forwarded to nodes that are better positioned to deliver them further to distant nodes. The aim of multihop routing is to elevate the transmission range of the broadcasted message in vehicular scenario. However, for the multihop forwarding strategy to be effective, traffic needs to be dense enough so that better positioned nodes exist within communication range [4]. We discuss each of these phases in the following sub-sections. Here, we outline the algorithm of the MMCS-MAC protocol in Algorithm 1.

Algorithm 1	MMCS-MAC protocol
Step 1	hello message signalling
Step 2	bandwidth division into frames
Step 3	mobility based cluster formation
Step 4	priority wise frame assignment to clusters in decreasing order of mobility
Step 5	safety message generation at node i
Step 6	message broadcasted to $(i+h_c)$ hop distance clusters. (Initially, the value of h_c is assumed to be 1)
Step 7	check for h_c
Step 8	broadcast the message
Step 9	increment h_c
Step 10	if $h_c \leq 4$, route the message to $(i+h_c)$ hop distance clusters
Step 11	goto step 7 and repeat the loop till $h_c = 4$
Step 12	if $h_c > 4$
Step 13	discard the message
Step 14	endif
Step 15	endif

A. CLUSTERING MECHANISM

The proposed scheme harnesses clustering based topology for safety message dissemination process. Nodes are clustered based upon their mobility. Nodes having near about same average speed form a cluster. Each node within a cluster is connected by one-hop intra-cluster link and different clusters link to each other through multihop topology. Since, the clustering algorithm is mobility based it does not require additional messages other than the dissemination of node's status messages (HELLO message signaling). Therefore, when nodes are on the road for the first time, they start sending their status messages without an elected CH. Once these messages are received by all nodes in the network range, they form a cluster following each other's mobility pattern. Cluster having maximum average speed is given highest priority to disseminate the safety message, which is implemented using the TDMA mechanism. We discuss this procedure here and outline it in Algorithm 2.

Each node in the network maintains positioning information by broadcasting HELLO messages. The HELLO message includes preliminary information such as node ID, position, mobility range etc. The HELLO broadcast period is defined as T_{HELLO}. When any node Y receives any other node X's HELLO message, Y will first check its similarity with X. A node will only consider neighbors moving in the same direction, and ignores broadcasts from traffic in the opposite direction. By means of broadcasting HELLO messages, each cluster records the neighboring cluster's positional information. This information serves as input to the clustering algorithm.

Algorithm 2	HELLO beacon signaling
Step 1	Every T_{HELLO}, X broadcasts HELLO beacons
Step 2	Each receiving neighbor checks for the similarity with X
Step 3	If true, Y calculates the coordinates of X
Step 4	X adds and updates its neighbor entry list

Cluster-based routing protocols involve four stages; CH selection, cluster formation, data aggregation and data communication [5]. In figure 2, nodes are shown to be clustered based upon their mobility. This clustering scheme negates the CH election overhead and produces relatively stable clustering structure. A cluster having longer travel duration (low mobility) has lower eligibility value to access the channel. Similarly, a cluster having shorter travel duration (high mobility) has higher eligibility value to access the channel. However, along with the inclusion of the speed difference, we need to know how to partition the network into minimum number of clusters such that when they are finally formed, the distribution of the nodes among them based on their mobility patterns is achieved with high probability. The proposed design employs a clustering approach whereby each cluster itself manages the intra-cluster communication using a TDMA scheme slot allocation. It is done by specifying when a node can transmit a message according to the availability of the slots in the cluster. Clusters are formed by nodes travelling in the same direction (one way). Therefore, all neighboring nodes used in our analysis are limited to those travelling in the same direction. However, the speed levels among them vary and this variation might be very high; thus, all neighboring nodes may not be suitable to be included in a cluster.

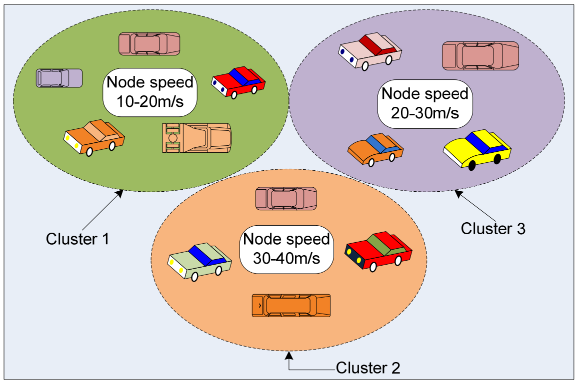


Figure 2. Mobility based clustering of nodes

The proposed algorithm results in maximum bandwidth utilization and minimum interference when the nodes are uniformly distributed on the roads. However, if the vehicular density is non-uniform, the slot requirement will differ, leading to interference, broadcast storming and inefficient bandwidth utilization. For that reason, we have assumed to have unvarying vehicular density, and that the nodes are moving with uniform speed for the duration of the simulation. The simulation time of 150 sec makes this assumption realistic and justifies the approach. The advantage of making such assumptions is to get rid of the overhead that the CH election process carries.

B. *TDMA TIME SLOT ASSIGNMENT*

The logic behind employing TDMA scheme is that it leverages contention less channel access by allocating time slots in one-hop radius distance to every cluster. However, when the destination of the message is several hops, the CM has to wait till its transmission slot arrives. We eliminate this delay by prioritizing slot assignment to clusters in decreasing order of their mobility. The slot assignment process assumes that each node may forward messages only to its one-hop neighbor, in the direction opposite to the direction of the node movement.

The proposed scheme rules out the implementation of channel switching during the synchronization interval as described in the legacy IEEE 1609.4 WAVE standard. A message can be delivered to any of the channels, irrespective of control channel interval (CCHI) and service channel interval (SCHI). More than that, entire DSRC bandwidth (75 MHz) is divided into frames. In order to enable multihop broadcast with minimal delay, every cluster is assigned a frame. Each frame is further divided into a number of slots. Frames are allocated to the clusters in a prioritized manner, assigning priority to the cluster having highest mobility. The cluster with maximum speed will have less time to access the channel. It has to be given higher priority with respect to other clusters. Similarly, based upon the mobility of clusters, different numbers of frames are assigned to them. Higher is the mobility, more numbers of frames are assigned. Figure 3 depicts the above discussed clustering based slot assignment process.

Since the proposed scheme follows a distributed approach, at the beginning of every TDMA frame the node randomly selects a transmission slot to transmit. All slots are equally likely to be selected. Each TDMA frame comprises 20 slots, each of 1 millisecond duration, so as to make it comparable to WAVE's CCH. Any event-driven message can be assigned to any of the slots to immediately broadcast the safety message. The nodes deliver the messages and vacate the slot within the frame assigned to them for the next messages from other nodes. Evidently, each node within the network becomes aware of the unallocated slots in the frame which gives them the opportunity to assign the slots amongst themselves. The number of frames per cluster is determined by the clustering algorithm during the clustering process and is given as input to the MAC layer. The slots are assigned to nodes in such a way that when a node receives a message travelling in a certain direction, it would immediately be able to forward the message to its next hop in the same direction [16].

In order to design a framework for intra and inter cluster communication in the proposed MMCS protocol, we need to design time slots in TDMA frames.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

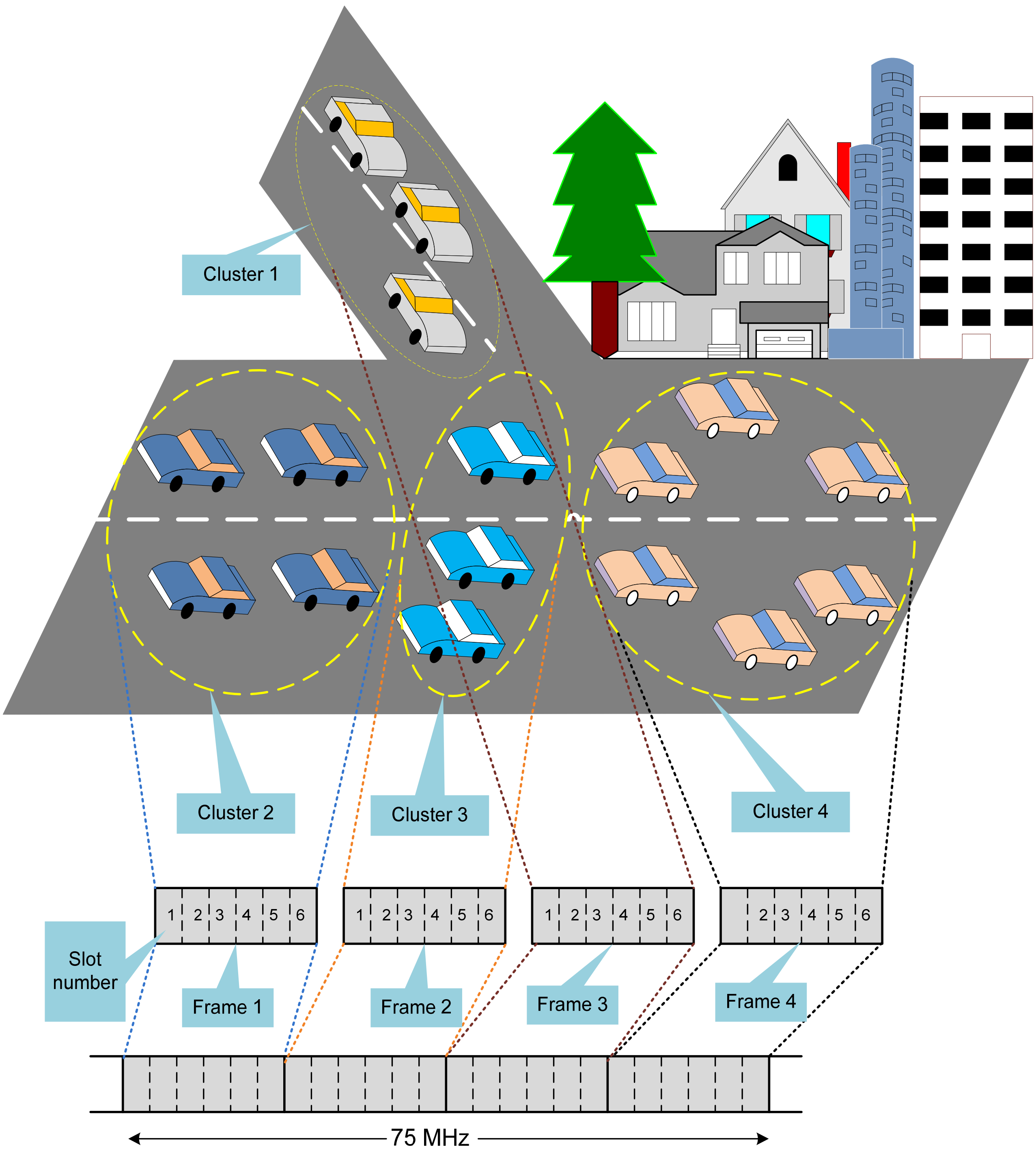


Figure 3. TDMA slot assignment based on clustering

As shown in figure 4, a TDMA frame consists of *n* time slots (slot 0 to slot n-1). Slot 0 is used to synchronize the first TDMA frame with the start of slot 1. Secondly, it broadcasts the slot-assignment state (SAS) within the cluster so that every node has a designated time slot for transmitting data. Slot 1 to slot n-1 of the TDMA frames are designated time slots used for data transmission.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

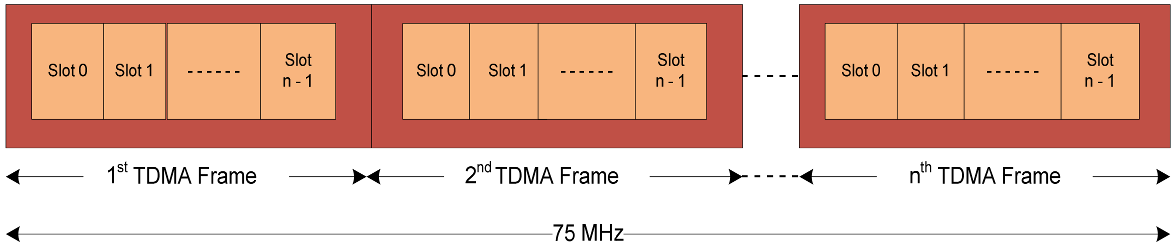


Figure 4. TDMA frame format

C. MULTIHOP MESSAGE ROUTING

In multihop message delivery, major issue lies in selecting the next hop path for data routing. The next hop is chosen from the nodes that lie in the direction of the destination node. This increases the probability of finding the shortest route [6]. For simplicity, we assume that there are n nodes in a network, and the position of any cluster c_1 in the network is Xc_1. From equation (1) it can be proved that any two clusters c_1 and c_2 are within the RF range of each other at any timestamp t if they satisfy the condition

$$\frac{\dfrac{Pn_1(t)}{(Xc_1(t)-Xc_2(t))^2}}{\dfrac{\sum Pc_1(t)}{(Xc_1(t)-Xc_2(t))^2}} > SINR \qquad (1)$$

where $Pn_1(t)$ is the transmit power of node n_1; $Xc_1(t) - Xc_2(t)$ is the distance between c_1 and c_2; SINR is the signal-to-interference-plus-noise ratio; and $\sum P_{c_1}(t)$ is the average transmit power of c_1.

Using equation (2), it can be shown that c_1 and c_2 are connected at time t if the distance between them is smaller than the transmission range T_r. That is,

$$\left(X_{c1}(t) - X_{c2}(t)\right) < T_r \qquad (2)$$

In multihop scenario, the messages can be quickly broadcasted among the connected nodes through a message routing scheme in which a node in a cluster broadcasts a packet to its neighboring clusters and each cluster that successfully receives the packet, rebroadcasts it to its immediate neighboring cluster. To make sure the messages transmit efficiently and correctly, the routing method in the multihop and dynamic topology network is very important [9].

Algorithm 3	Multihop message routing
Step 1	start routing
Step 2	while true
Step 3	if REQ received
Step 4	get the source ID and h_c
Step 5	endif
Step 6	if the Tx and Rx outside cluster
Step 7	update SAS
Step 8	endif
Step 9	rebroadcast REQ
Step 10	$h_c = h_c + 1$
Step 11	rebroadcast REQ
Step 12	if $h_c > N$
Step 13	discard the REQ
Step 14	end if
Step 15	endofwhile

Figure 5 represents the flow diagram of the proposed MMCS-MAC protocol. Initially, HELLO message signaling allows all the nodes in the network to get acquainted with each other's coordinates. Secondly, we divide the entire DSRC band into frames so that full bandwidth remains available for safety message transmission. Next, based upon the mobility pattern gathered to HELLO message beaconing, clusters are formed. Moving ahead, priority wise frames are assigned to the clusters in decreasing order of their mobility. That is, cluster with maximum speed is assigned more number of frames to transmit. This not only ensures reliability but leverages better channel utilization as well. These frames are further divided into a number of slots. Each node is assigned a slot to transmit its message. Now, let us assume that a safety message is generated at node i. This message is broadcasted to (i+h_c) hop distance clusters where h_c is the hop-count of the message. Initially, the value of h_c is assumed to be 1. When the message is received by one-hop distant cluster, it checks for the current h_c. If h_c < N, where N=4, it broadcasts the message and increments the h_c by 1. Likewise, the message is broadcasted and relayed up to N hops distant clusters. We take h_c to be 4 because when h_c becomes greater

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

than N hops, the message is no longer relayed and is discarded as obtained from the simulation results.

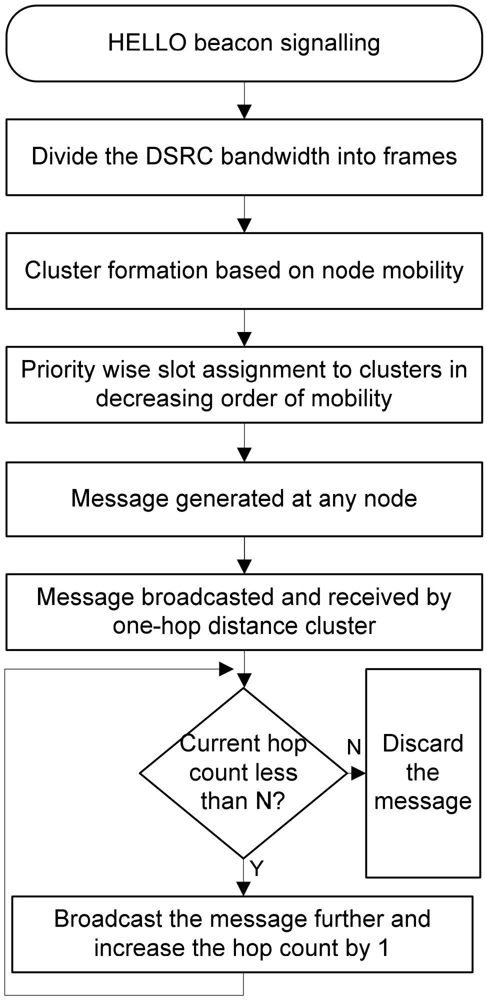


Figure 5. Flow diagram of MMCS-MAC

V. PERFORMANCE EVALUATION

In this section, we evaluate and compare the performance metrics of the proposed protocol with (i) IEEE 802.11p standard and (ii) other existing protocols such as Distributed Multichannel Mobility-Aware Cluster-Based MAC protocol (DMMAC) [8], Cluster-Based Beacon Dissemination Process (CB-BDP) [11], and WAVE-enhanced Service message Delivery (WSD) [18] with respect to the vehicular density. We show that clustering reduces delay in multihop broadcasting scenario. The results also raise concerns about the existing standard's capability of providing safety at the road level, and thus justify the need for protocol enhancements that take into account the QoS requirements of vehicular applications. Due to the impact of relative speed in V2V communication, an effective MAC protocol should provide priority to a node with higher mobility to transmit before it moves out of the communication range.

A. SIMULATION SETUP

The simulations are carried out for a 4-lane highway with nodes moving in both directions. Node speed varies between 10 to 40 m/s. All nodes have the same IEEE 802.11p standard MAC parameters for V2V communication in multihop ad-hoc region. The simulation time is set to 150s, and the transmission range of each node is up to 300 m. The message size is arbitrarily taken to be 512 bytes which is transmitted at the rate of 6 Mbps since it is the prescribed data rate for DSRC safety applications [25]. The data transfer rate and ad-hoc coverage range is taken as per the IEEE 802.11p standard. Vehicular density is assumed to be uniform and the number of nodes contending for the channel varies from 5 to 40, in steps of 5. For the sake of a diversified comparison, the proposed MMCS-MAC protocol is compared with the related performance metrics of various existing protocols since they carry near resemblance to this work. Table 1 summarizes the parameters used in our simulation. The parameters are taken to model a simplified, yet realistic vehicular traffic scenario on highways.

TABLE 1.Simulation parameters

Parameter	**Values**
Number of nodes	5 - 40
Node's speed	10 - 40 m/s
Simulation area	10000 x10000 (m^2)
Simulation time	150 sec
Data rate	6 Mbps
Number of lanes	4
Scenario	Highway
Transmission range	300 m
Interface queue type	Queue/DSRC
Interface queue length	50
Network interface	Phy/WirelessPhyExt
MAC interface	802. 11Ext
Message size	512 Bytes
Propagation model	Two Ray Ground
Modulation type	BPSK
Antenna type	Omni Antenna

B. PERFORMANCE METRICS

In order to evaluate the proposed protocol's performance for safety message dissemination, following metrics are defined:

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

i) Packet delivery ratio (PDR): it measures the success ratio of the transmissions, that is, the ratio of the number of packets successfully received to the total number of packets sent. PDR is analyzed as

$$PDR = \frac{\sum_{i=1}^{n} x_i}{\eta_{T_{RF}} \sum_{i=1}^{n} y_i} \tag{3}$$

where x_i is the number of packets received by node i, y_i is the number of packets transmitted by node i and $\eta_{T_{RF}}$ is the average number of neighboring nodes in the RF transmission range. The value of $\eta_{T_{RF}}$ is approximated using the vehicular density. Figure 6 shows the comparison among the proposed protocol, CB-BDP and IEEE 802.11p standard for the PDR with respect to the vehicular density. As the number of nodes increase, the PDR tends to increase because the probability of more packets getting delivered rises. The proposed MMCS-MAC protocol is seen to perform better when compared to the other two protocols for this metric. This is because the proposed protocol attempts to transfer messages up to four hops, thereby enhancing the probability of message reception.

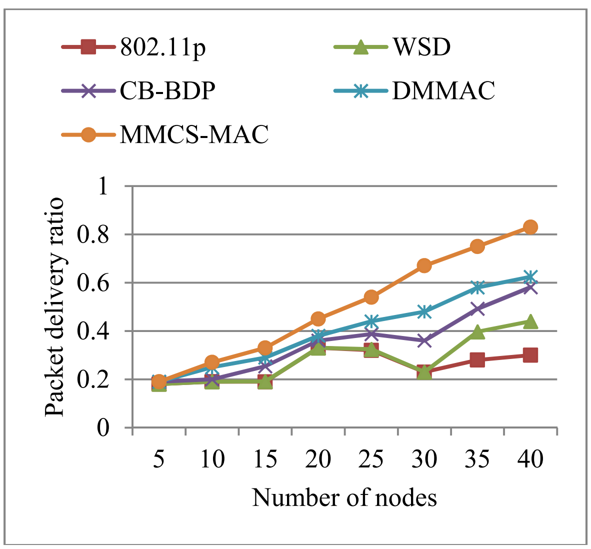


Figure 6. Comparison of PDR

ii) Throughput (δ): defined as the rate of successful data delivery in the network per unit time. This metric gives the measure of the how much data is received in the network. It is averaged per node and analytically defined as

$$\delta = \frac{\sum_{i=1}^{n} x_i}{T_s n} \tag{4}$$

where x_i is the number of packets received by node i, T_s is the simulation time in seconds, and n is the total number of nodes in the network and. Figure 7 shows the throughput range attained by different protocols under study. Clearly, MMCS-MAC outperforms the other two protocols. It attributes to the TDMA based clustering whereby the nodes within a cluster self-assign the slots to disseminate the safety messages, avoiding the cluster maintenance overhead. This results in higher rate of successful message reception. For all the three protocols, throughput rises till 25 nodes. However, beyond this range it surges between the range of 30-35 nodes and rises again as the vehicular density increases.

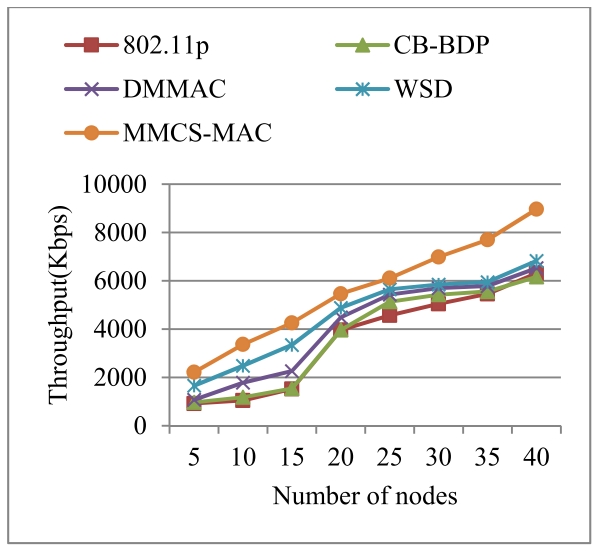


Figure 7. Comparison of Throughput

iii) Packet loss ratio (P_{LR}): data packets fail to reach their destination and are lost during transmission. Major cause of packet loss is typically network congestion. Packet loss is measured as a ratio of number of packets lost with respect to total packets transmitted. It can be formulated as

$$P_{LR} = \frac{\text{number of packets lost}}{\text{total number of packets transmitted}}$$

Figure 8 shows the P_{LR} for different protocols. Whereas the MMCS slightly performs better than CB-BDP, the performance of IEEE 802.11p degrades drastically as the vehicular density increases.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

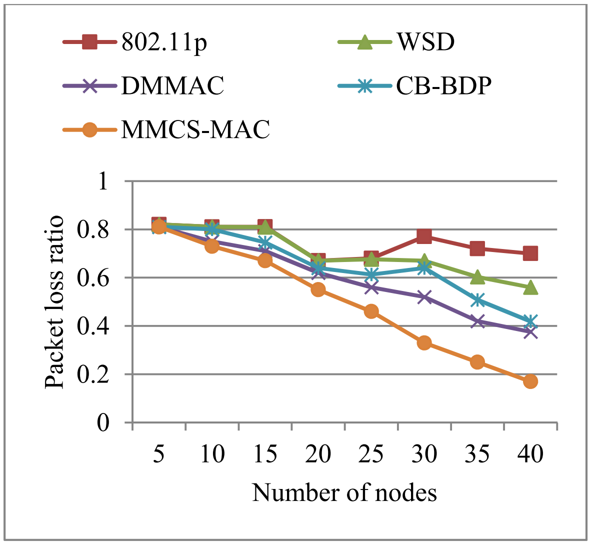


Figure 8. Comparison of Packet Loss Ratio

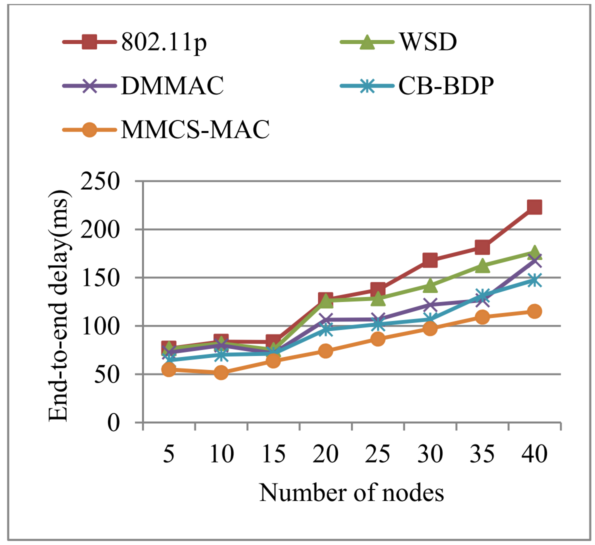


Figure 9. Comparison of Average End-to-End Delay

This attests to the fact that the DCF of the standard protocol is not suitable for safety message dissemination under highly dense vehicular scenario. The reason for the improved performance of the proposed protocol relates again, to the high probability of message reception.

iv) Average end-to-end delay (t_{avg}): time elapsed during sending a packet from the source node and reception of the packet at the destination node is the end-to-end delay of that packet. The total delay of all delivered packets divided by the total number of packets delivered gives the average end-to-end delay of the network. It is given by

$$t_{avg} = \frac{\sum_{i=1}^{n} t_i}{n} \quad (5)$$

Figure 9 evaluates and compares the proposed protocol with the WSD and IEEE 802.11p standard. We introduce WSD scheme here as it recognizes delivery delay as a stringent QoS requirement as far as safety-related applications are concerned. It is observed that as the number of nodes increase, the delay rises. This is pretty obvious owing to the multihop scenario where each intermediate cluster follows a protocol so as to route the message to its one-hop distance cluster. The delay encountered in the MMCS-MAC protocol is comparable to that of WSD, perhaps with slight improvement being seen over the latter. This improvement is attributed to the fact that the proposed protocol focuses on multihop dissemination, unlike to WSD scheme that targets single hop dissemination.

v) Probability of successful delivery (PSD): a high level of certainty is required while delivering safety messages. It not only relates to the reliable data delivery but also with the overall efficiency of the network. Figure 10 compares the MMCS-MAC with WSD and 802.11 p MAC protocols. Whereas the standard protocol doesn't show credible performance, WSD demonstrates better results. Notwithstanding, MMCS-MAC shows high probability of successful message delivery. However, all protocols show a decreasing trend with increasing vehicular density. For the proposed protocol, the probability lies in the range of 70% to 95% for low density (up to 20 nodes). As the number of nodes increase, the probability decreases. This shows that as the number of hops increase, the certainty of a message getting delivered falls.

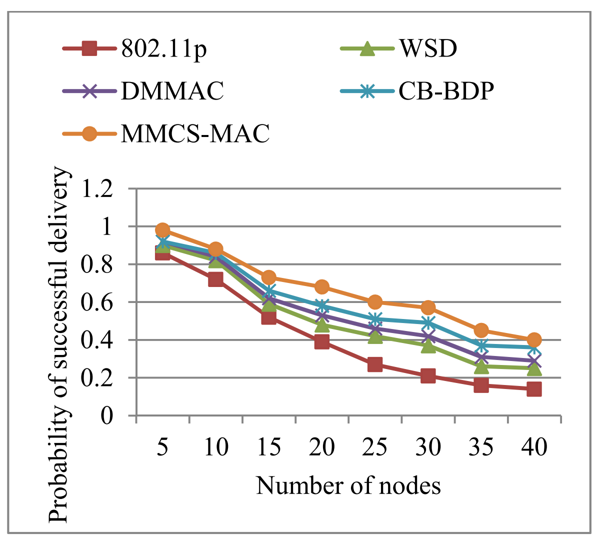


Figure 10. Comparison of PSD

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

vi) Reliability: probability that a cluster and its one-hop distant neighboring cluster will transmit and receive the message successfully. Reliability is one of the most important parameter when we focus on safety message dissemination. In figure 11, MMCS-MAC is compared with DMMAC and the standard protocol. DMMAC protocol focuses on reliability in delivering safety messages under similar vehicular scenario. It can be seen that both DMMAC and MMCS-MAC performs consistently well under the specified simulation conditions. This justifies the reason for comparison with DMMAC. The system's reliability is seen to be high in low-density network, and slightly decreases as the vehicular density increases.

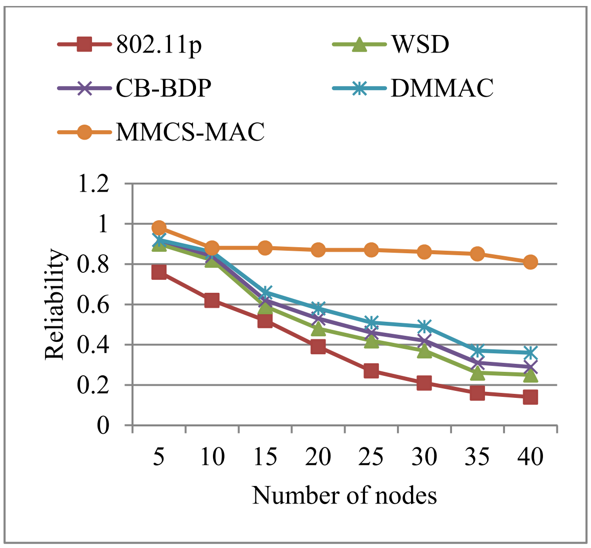


Figure 11. Comparison of Reliability

This is possibly due to the increasing number of hops which tends to increase with increasing density. However, the standard protocol fails to demonstrate high level of reliable transmission with increasing vehicular density which again questions its applicability to cater to dissemination of safety messages.

vii) Safety message travel time: the time taken by a safety message sent by a node to reach its one-hop distance neighbor. In figure 12 it is shown that as the number of nodes increase, the travel time decreases. It so happens because with increasing node density, hopping will increase, resulting in faster message delivery. This result goes in favor of the fact that more is the number of clusters, better will be the multihop broadcasting. Moreover, the decrease in the node density results in increasing the safety message travel time since nodes may struggle to find a neighboring node to carry the message forward. MMCS-MAC is seen to perform better than the other two protocols because of two reasons. Firstly, MMCS-MAC does not require a cluster-head selection. This reduces the additional time that would have been consumed in the process. Secondly, since the vehicular density is assumed to be constant for the simulation duration, HELLO message beaconing is performed only once, when the nodes are on the road for the first time. This further helps in reducing the travel time.

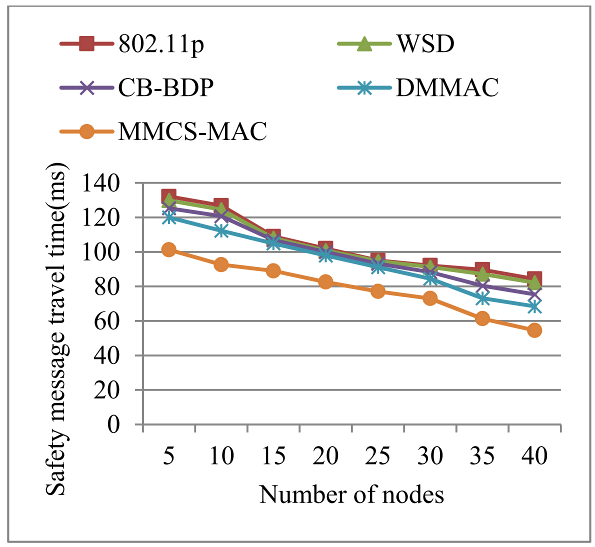


Figure 12. Comparison of Message Travel Time

viii) Packet inter-reception time (PIRT): defined as the time elapsed between the receptions of two successive beacons at any specific node. Evaluating the PIRT is justified by the observation that it is an important beaconing metric, as well as an important class of active safety applications, such as collision warning, emergency braking and transit node signaling etc. These applications mandate its requirement in terms of maximum tolerable PIRT [26]. In figure 13, we compare MMCS-MAC with the DCF of the IEEE 802.11p standard. The reason for comparing the proposed protocol only with the standard protocol is that till now, no such MAC scheme showing resemblance to the proposed protocol has evaluated this metric and hence comparison could not be made.

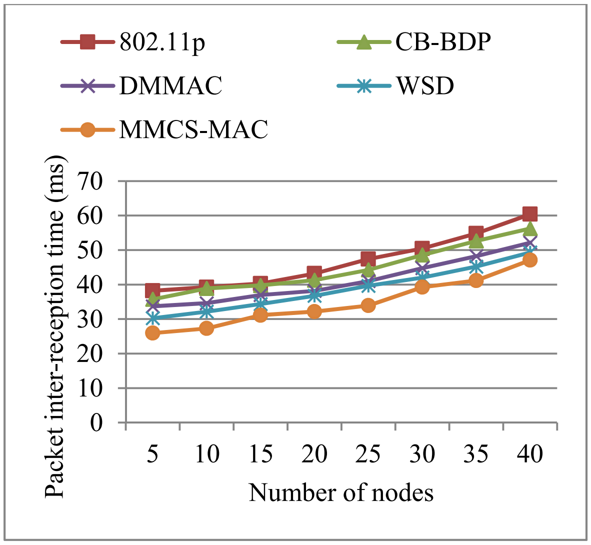


Figure 13. Comparison of PIRT

From the obtained results upon evaluation, it is observed that with increasing node density, PIRT increases for both the protocols. Moreover, PIRT for MMCS-MAC is lower than the legacy standard which is a desirable observation. This improvement over the IEEE 802.11p standard protocol is due to the adoption of mobility based clustering scheme that leverages faster transmission and reception of safety messages. This attests that the proposed protocol performs better under dense vehicular scenario.

VI. CONCLUSION

This paper presented a novel mobility dependent MAC protocol suitable for traffic safety applications in VANET. The aim was to define a scheme that is able to scale over a number of nodes and deliver the messages in real-time scenario. The protocol harnesses the clustering based TDMA scheme for multihop message dissemination in inter-vehicular communication which is fully distributed and does not require a cluster-head selection. Clustering the nodes based on their speed increases the stability. The decision of not electing the CH helps in reducing the overhead of cluster maintenance. The scheme leverages on the fact that real-time traffic having higher-sensitivity should gain more priority to acquire time slots than non-real-time traffic with lower-priority. The TDMA mechanism allocates time slots to the clusters based on their mobility pattern. Frame synchronization between different clusters allows the protocol to ensure reliable and timely delivery of safety messages. We show how it could significantly improve the efficiency and PDR. Multihop routing has been accomplished for up to four hops. Simulations have been performed using NS-2.34 network simulator. From the simulation results, it is observed that with the formation of small-sized cluster, the network spends less time than IEEE 802.11p. Moreover, it attests to the fact that the existing WAVE standard succumbs to perform under high vehicular density and is incapable to ensure reliable dissemination of safety messages. Additionally, when the node density is increased, the protocol takes less waiting time before a node can effectively transmit data.

From comparison with other related works, it can be clearly stated that the performance of MMCS-MAC outshines not only the performance of the IEEE 802.11p standard but also of other protocols it is compared with in delivering safety messages to the intended recipients. The designed algorithm helps MMCS-MAC to maintain a high level of reliability, particularly under high vehicular density along with assuring high packet delivery ratio and throughput.

A direction for future work could be to extend the proposed scheme for different traffic types based on the assigned priorities to them; provisioning of non-safety messages; and integrating the concept of adaptive contention window while allocating slots.

REFERENCES

[1] Rawashdeh ZY, Mahmud SM (2008) Media access technique for cluster-based vehicular ad hoc networks. In: 68th IEEE Vehicular Technology Conference, VTC 2008-Fall, Calgary, Canada, (pp. 1-5)

[2] Sheu TL, Lin YH (2014) A Cluster-based TDMA System for Inter-Vehicle Communications. Journal of Information Science and Engineering 30(1):213-231

[3] Hassan MI, Vu HL, Sakurai T (2011). Performance analysis of the IEEE 802.11 MAC protocol for DSRC safety applications. IEEE Trans Veh Technol 60(8):3882-3896

[4] Lee KC, Lee U, Gerla M (2010). Survey of routing protocols in vehicular ad hoc networks. Advances in vehicular ad-hoc networks: Developments and challenges:149-170

[5] Skordylis A, Trigoni N (2008) Delay-bounded routing in vehicular ad-hoc networks. In: 9th ACM international symposium on Mobile ad hoc networking and computing (MobiHoc), Hong Kong SAR, China, (pp. 341–350)

[6] Mu'azu AA, Jung LT, Lawal I, Shah PA (2013) A QoS approach for cluster-based routing in VANETS using TDMA scheme. In: IEEE International Conference on ICT Convergence (ICTC), Jeju, Korea, (pp. 212-217)

[7] Willke TL, Tientrakool P, & Maxemchuk NF (2009). A survey of inter-vehicle communication protocols and their applications. IEEE Communications Surveys & Tutorials 11(2):3-20

[8] Hafeez KA, Zhao L, Mark JW, Shen X, Niu Z (2013). Distributed multichannel and mobility-aware cluster-based MAC protocol for vehicular ad hoc networks. IEEE Trans Veh Technol 62(8):3886-3902

[9] Tian D, Wang Y, Xia H, Cai F (2013). Clustering multi-hop information dissemination method in vehicular ad hoc networks. IET Intelligent Transport Systems 7(4):464-472

[10] Van Tien P, Viet BD, Dzung NT, Cuong TH (2012) Location-demanded data delivery over vehicular ad hoc networks. In: 4th IEEE International Conference on

Communications and Electronics (ICCE), Hue, Vietnam, (pp. 108-113)

[11] Allouche Y, Segal M (2015). Cluster-Based Beaconing Process for VANET. Vehicular Communications 2(2):80-94

[12] Almalag MS, Olariu S, Weigle MC (2012) Tdma cluster-based mac for vanets (tc-mac). In: IEEE International Symposium on World of Wireless, Mobile and Multimedia Networks (WoWMoM),San Francisco, USA, (pp. 1-6)

[13] Shea C, Hassanabadi B, Valaee S (2009) Mobility-based clustering in VANETs using affinity propagation. In: IEEE Global Telecommunications Conference (GLOBECOM 2009), Honolulu, USA, (pp. 1-6)

[14] Booysen MJ, Zeadally S, Van Rooyen, GJ (2013) Impact of neighbor awareness at the MAC layer in a Vehicular Ad-Hoc Network (VANET). In: 5th IEEE International Symposium on Wireless Vehicular Communications (WiVeC), Dresden, Germany, (pp. 1-5)

[15] Zhang Z, Boukerche A, Pazzi R (2011) A novel multi-hop clustering scheme for vehicular ad-hoc networks. In: 9th ACM international symposium on Mobility management and wireless access, Florida, USA, (pp. 19-26)

[16] Zelikman D, Segal M (2014). Reducing Interferences in VANETs. IEEE Transactions on Intelligent Transportation Systems 16(3):1582–1587

[17] Mammu K, Sidhik A, Hernandez-Jayo U, Sainz N (2013) Cluster-based MAC in VANETs for safety applications. In: IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI), Mysore, India, (pp. 1424-1429)

[18] Ghandour AJ, Di Felice M, Artail H, Bononi L (2014). Dissemination of safety messages in IEEE 802.11 p/WAVE vehicular network: Analytical study and protocol enhancements. Pervasive and Mobile Computing 11:3-18

[19] Najafzadeh S, Ithnin N, Razak SA, Karimi R (2014). BSM: broadcasting of safety messages in vehicular ad hoc networks. Arabian Journal for Science and Engineering 39(2):777-782

[20] Omar H, Zhuang W, Li L (2013). VeMAC: A TDMA-based MAC protocol for reliable broadcast in VANETs. IEEE Transactions on Mobile Computing 12(9):1724-1736

[21] Bharati S, Zhuang W (2013). CAH-MAC: cooperative ADHOC MAC for vehicular networks. IEEE Journal on Selected Areas in Communications 31(9):470-479

[22] Liu J, Ren F, Miao L, Lin C (2011). A-ADHOC: An adaptive real-time distributed MAC protocol for Vehicular Ad Hoc Networks, Mobile Networks and Applications 16(5): 576-585

[23] Gupta N, Prakash A, Tripathi R (2015). Medium access control protocols for safety applications in Vehicular Ad-Hoc Network: A classification and comprehensive survey, Vehicular Communications 2(4):223-237

[24] Haas JJ, Hu YC (2010) Communication requirements for crash avoidance. In: 7th ACM international workshop on VehiculAr InterNETworking, Chicago, USA, (pp. 1-10)

[25] Chen Q, Jiang D, Taliwal V, Delgrossi L (2006) IEEE 802.11 based vehicular communication simulation design for NS-2. In: 3rd ACM international workshop on Vehicular ad hoc networks, Los Angeles, USA, (pp. 50-56)

[26] Sepulcre M, Gozalvez J (2012) Experimental evaluation of cooperative active safety applications based on V2V communications. In 9th ACM international workshop on Vehicular inter-networking, systems, and applications, Lake District, UK, (pp. 13-20)

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 3, March 2016

Clustering of Hub and Authority Web Documents for Information Retrieval

Kavita Kanathey
Computer Science
Barkatullah University
Bhopal,MP,India

R. S. Thakur
Department of Computer Application
Maulana Azad National Institute of Technology (MANIT)
Bhopal, MP, India

Shailesh Jaloree
Department of applied Mathematics
SATI,Vidisha,Bhopal,MP,India

Abstract- Due to the exponential growth of World Wide Web (or simply the Web), finding and ranking of relevant web documents has become an extremely challenging task. When a user tries to retrieve relevant information of high quality from the Web, then ranking of search results of a user query plays an important role. Ranking provides an ordered list of web documents so that users can easily navigate through the search results and find the information content as per their need. In order to rank these web documents, a lot of ranking algorithms (PageRank, HITS, Weight PageRank) have been proposed based upon many factors like citations analysis, content similarity, annotations etc. However, the ranking mechanism of these algorithms gives user with a set of non classified web documents according to their query. In this paper, we propose a link-based clustering approach to cluster search results returned from link based web search engine. By filtering some irrelevant pages, our approach classified relevant web pages into most relevant, relevant and irrelevant groups to facilitate users' accessing and browsing. In order to increase relevancy accuracy, K-mean clustering algorithm is used. Preliminary evaluations are conducted to examine its effectiveness. The results show that clustering on web search results through link analysis is promising. This paper also outlines various page ranking algorithms.

Keywords - World Wide Web, search engine, information retrieval, Pagerank, HITS, Weighted Pagerank, link analysis.

I. INTRODUCTION

The World Wide Web is a famous and interactive way to disseminate information nowadays. The Web is the largest information repository for knowledge reference. The web is huge, semi-structured, dynamic, and heterogeneous and broadly distributed global information service center [5]. Finding relevant web pages of highest quality to the users based on their queries becomes increasingly difficulty. This can be observed by the researcher that most of the web documents collected by web spider are not relevant to the query of the user. It makes in-convenience for the user to filter out irrelevant information from these search results, hence leading to waste of time. For these reasons, the cluster search engine provides a way to find the information, by returning a set of classified web pages.

An important class of search engine that offer search results based on hypertext links between sites can be termed as Link Based Search Engine. Rather than providing results based on keywords or the content of the web documents, sites are ranked based on the quality and quantity of other web sites linked to them. In this system, user submits a query to the meta-search engine. The meta-search engine searches for the relevant results of users query. From the set of results retrieved from web search engine, they are formed as a meta-directory tree. This tree structure helps the user to retrieve information with high relevancy.

The relevancy of web page can be obtained by considering the number of in-links and out-links present in a particular web page. When the web page has more number of out-links to a relevant page, then that page can be considered as a central page. From this central page, all other web pages are compared for similarity and the most similar pages are grouped together. The grouping of most similar pages together is known as clustering. Clustering can be done based on different algorithms such as hierarchical, k-means, partitioning, etc.

The simplest unsupervised learning algorithm that solve clustering problem is K- Means algorithm. It is a simple and easy way to classify a given data set through a certain number of clusters.

When the documents are clustered [9] using K-Means algorithm, the cluster contains more similar documents and it increases the relevancy rate of search results. When a user requests for a query after these clustering process, they get only the most relevant cluster which matches the request. They will not get any of the irrelevant pages. So, it increases the efficiency of search results and reduces computational time and search space.

The paper is organized as follows. Section II is an assessment of previous related works of link analysis and clustering in web domain. In Section III**,** we describe the existing system. Subsequently in Section IV we describe our proposed approach in detail. In Section V, We conclude our paper with some discussions.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

II. RELATED WORK

In order to retrieve more relevant documents, various Link Analysis Algorithms have been proposed. Three important algorithms PageRank,Weight PageRank and hypertext Induced Topic Search(HITS) are discussed below in detail and compared

A. PageRank Algorithm

The PageRank [1] is the link analysis algorithm that was developed by S. Brin and L. Page during their Ph.D. at Stanford University based on the citation analysis. This algorithm is used by the famous search engine GOOGLE. PageRank algorithm applied the citation analysis in web search by treating the incoming links as citations to the web pages. This algorithm is based on the concepts that if a page contains “important” links towards it then the links of this page towards the other page are also to be considered as “important” pages. The PageRank considers the back link in deciding the rank score. If the addition of the all the ranks of the back links is large then the page then it is provided a large rank. Therefore, PageRank provides a more advanced way to compute the importance or relevance of a web page than simply counting the number of pages that are linking to it. If a back link comes from an important page, then that back link is given a higher weighting than those back links comes from non-important pages. In a simple manner, link from one page to another page may be considered as a vote. However, not only the number of votes a page receives is considered important, but the importance or the relevance of the ones that cast these votes as well.

Assume any arbitrary page A has pages T_1 to T_n pointing to it (inlink). PageRank can be calculated by the following Eq. (1):

$$PR(A) = (1\text{-}d) + d[PR(T_1)/C(T_1) + \ldots + PR(T_n)/C(T_n)] \ldots\ldots (1)$$

Where PR(A) is the PageRank of page A; $PR(T_i)$,for i=1…n, is the PageRank of page T_i which links to page A,C $((T_i)$; for i=1…n, is the outbound links on page T_i, and d is a damping factor, usually sets it to 0.85.
Consider a small web consisting of three web pages P, Q and R as shown in fig.1

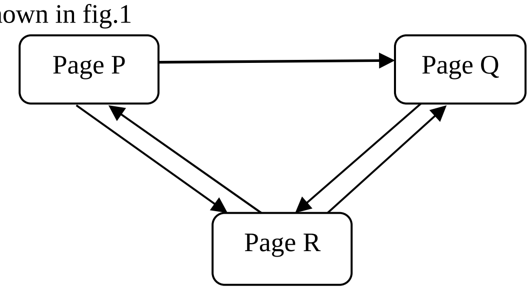


Figure.1 Hyperlink Structure of web pages

The PageRank for pages P, Q and R are calculated manually by using Eq. (1). Let us assume the initial PageRank as 1.0 and do the calculation. The damping factor d is set to 0.85:

PR (P) = (1-d) + d [PR(R)/C(R)]
= (1-0.85) +0.85(1/2)
= 0.15+0.425
= 0.575 (1a)

PR(Q)= (1-d) +d [PR (P)/C(P)+ PR(R)/C(R)]
= (1-0.85)+0.85[0.575/2+1/2]
=0.819 (1b)

PR(R) = (1-d) +d [PR (P)/C (P) + PR (Q)/C (Q)]
= (1-0.85)+0.85[0.575/2+0.819/1]
= 1.091 (1c)

Do the second iteration by taking the above PageRank value from (1a), (1b) and (1c):

PR (P) = (1-d) + d [PR(R)/C(R)]
= 0.15+0.85[1.091/2]
= 0.614 (2a)

PR (Q) = (1-d) +d [PR (P)/C (P) + PR(R)/C(R)]
= 0.15+0.85[0.614/2+1.091/2]
=0.875 (2b)

PR(R) = (1-d) +d [PR (P)/C (P) + PR (Q)/C (Q)]
= 0.15+0.85[0.614/2+0.875/1]
= 1.155 (2c)

Do the third iteration by taking the above PageRank values from (2a), (2b) and (2c):

PR (P) = (1-d) + d [PR(R)/C(R)]
= 0.15+0.85[1.155/2]
= 0.578 (3a)

PR (Q)= (1-d) +d [PR (P)/C (P) + PR(R)/C(R)]
=0.15+0.85[0.578/2+1.155/2]
=0.886 (3b)

PR(R) = (1-d)+d[PR(P)/C(P)+ PR(Q)/C(Q)]
= 0.15+0.85[0.578/2+0.886/1]
=1.148 (3c)

After doing many more iterations of the above calculation, the PageRanks arrived as shown in Table 1.
For a smaller set of pages, the computation is easier but for a Web having billions of pages; the above computation becomes more difficult. As shown in the Table 1, you can notice that PR(R) >PR (Q)>PR (P).So the link analysis becomes very important in the PageRank. From the Table 1, after the iteration 15, the PageRank for the pages gets normalized. The PageRank gets converged to a reasonable tolerance.

Table 1: Iterative calculation for PageRank

Iteration	*PR(P)*	*PR(Q)*	*PR(R)*
0	1.000	1.000	1.000
1	0.575	0.819	1.091
2	0.614	0.875	1.155
3	0.578	0.886	1.148
………	……….	………..	……….
15	0.701	0.999	1.297
16	0.701	0.999	1.297

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

B. Weighted Page Rank Algorithm

Weighted PageRank Algorithm [4] was proposed by Wenpu Xing and Ali Ghorbani. Weighted PageRank algorithm (WPR) is an extension of the original PageRank algorithm. This algorithm assigns larger rank values to more important (popular) pages instead of dividing the rank value of a page evenly among it's outlink pages. Each outlink page gets a value proportional to its popularity (its number of inlinks and outlinks). The popularity from the number of inlinks and outlinks is recorded as $W^{in}_{(v,u)}$ and $W^{out}_{(v,u)}$, respectively.

$W^{in}_{(v,u)}$ is the weight of *link*(*v, u*) calculated based on the number of inlinks of page *u* and the number of inlinks of all reference pages of page *v*.

$$W^{in}_{(v,u)} = I_u / \Sigma_{p \in R(v)} Ip \quad (1)$$

Where I_u and I_p represent the number of inlinks of page *u* and page *p*, respectively. *R* (*v*) denotes the reference page list of page *v*.

$W^{out}_{(v,u)}$ is the weight of *link*(*v, u*) calculated based on the number of outlinks of page *u* and the number of outlinks of all reference pages of page *v*.

$$W^{out}_{(v,u)} = O_u / \Sigma_{p \in R(v)} O_p \quad (2)$$

Where O_u and O_p represent the number of outlinks of page *u* and page *p*, respectively. *R* (*v*) denotes the reference page list of page *v*.

Considering the importance of pages, the original PageRank formula is modified as

$$WPR(u) = (1 - d) + d \, \Sigma_{v \in B(u)} WPR(v) \, W^{in}_{(v,u)} \, W^{out}_{(v,u)} \quad (3)$$

Use the same hyperlink structure as shown in Fig. 1 and perform the WPR computation. The WPR equations for page P, Q and R are as follows.

$$WPR(P) = (1-d) + d\,[WPR(R) W^{in}_{(R,P)} \, W^{out}_{(R,P)}] \quad (1a)$$

$$WPR(Q) = (1-d) + d\,[WPR(P) W^{in}_{(P,Q)} \, W^{out}_{(P,Q)} + WPR(R) W^{in}_{(R,Q)} . W^{out}_{(R,Q)}] \quad (1b)$$

$$WPR(R) = (1-d) + d\,[WPR(P) . W^{in}_{(P,R)} W^{out}_{(P,R)} + WPR(Q) W^{in}_{(Q,R)} . W^{out}_{(Q,R)}] \quad (1c)$$

Let us assume the initial PageRank as 1.0 and do the calculation. The damping factor d is set to 0.85: The inlink and outlink weights are calculated as follows:

$$W^{in}_{(R,P)} = I_P / (I_P + I_Q) = 1/(1+2) = 1/3 \quad (1.1a)$$

$$W^{out}_{(R,P)} = O_P / (O_P + O_Q) = 2/2+1 = 2/3 \quad (1.1b)$$

By substituting the values of equation (1.1a) and (1.1b) in (1a), you will get the WPR for page P.

$$WPR(P) = 0.15 + 0.85[1*1/3*2/3] = 0.338 \quad (2a)$$

The inlink and outlink weights for page Q are calculated as follows:

$$W^{in}_{(P,Q)} = I_Q / (I_Q + I_R) = 2/2+2 = 1/2 \quad (2.1a)$$

$$W^{out}_{(P,Q)} = O_Q / (O_Q + O_R) = 1/1+2 = 1/3 \quad (2.1b)$$

$$W^{in}_{(R,Q)} = I_Q / (I_Q + I_P) = 2/2+1 = 2/3 \quad (2.1c)$$

$$W^{out}_{(R,Q)}] = O_Q / (O_Q + O_P) = 1/1+2 = 1/3 \quad (2.1d)$$

By substituting the values of equation (2.1a) ,(2.1b),(2.1c)and (2.1d) in (1b),You will get the WPR for page Q.

$$WPR(Q) = 0.15+0.85[0.338*1/2*1/3+1*2/3*1/3] = 0.386 \quad (2b)$$

The inlink and outlink weights for page R are calculated as follows:

$$W^{in}_{(P,R)} = I_R / (I_R + I_Q) = 2/2+2 = 1/2 \quad (3.1a)$$

$$W^{out}_{(P,R)} = O_R / (O_Q + O_R) = 2/2+1 = 2/3 \quad (3.1b)$$

$$W^{in}_{(Q,R)} = I_R / (I_R + I_P) = 2/2+1 = 2/3 \quad (3.1c)$$

$$W^{out}_{(Q,R)} = O_R / (O_P + O_R) = 2/2+2 = 1/2 \quad (3.1d)$$

By substituting these values in (1c), you will get the WPR for page R.

$$WPR(R) = 0.15+0.85[0.338*1/2*1/3+0.386*2/3*1/3] = 0.354$$

After doing many more iterations of the above calculation, the Weighted PageRanks arrived as shown in Table 2.

Table 2. Iterative calculation for PageRank

Iteration	WPR(P)	WPR(Q)	WPR(R)
0	1.000	1.000	1.000
1	0.338	0.386	0.354
2	0.217	0.248	0.282
3	0.203	0.232	0.273
4	0.201	0.231	0.272
5	0.201	0.230	0.272
6	0.201	0.230	0.272

As shown in table 2, WPR(R) >WPR (Q)>WPR (P) in less iteration.

C. Hypertext Induced Topic Search (HITS) Algorithm

The HITS algorithm is proposed by Kleinberg in 1999. Kleinberg identifies two different forms of Web pages called hubs and authorities. Authorities are pages having important contents. Hubs are pages that act as resource lists, guiding users to authorities. Thus, a good hub page for a subject points to many authoritative pages on that content and a good authority page is pointed by many good hub pages on the same subject. Hubs and Authorities and their calculations are shown in Fig. 2. Kleinberg says that a page may be a good hub and a good authority at the same time. This circular relationship leads to the definition of an iterative algorithm called Hyperlink Induced Topic Search (HITS) [6].

The HITS algorithm treats WWW as a directed graph G (V, E), where V is a set of vertices representing pages and E is a set of edges that correspond to links.

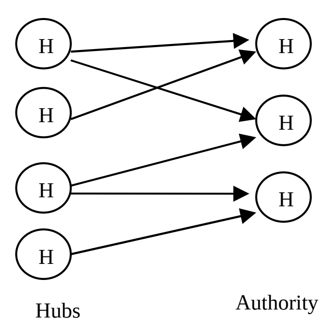


Figure.2 Hubs and Authorities

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Since a good authority is pointed to by many good hubs and a good hub points to many good authorities, such mutually reinforcing relationship can be represented as:

$$x_p = \Sigma_{q:(q;p) \in E}\, y_q \qquad (1)$$

$$y_p = \Sigma_{q:(q;p) \in E}\, x_q \qquad (2)$$

where x_p is the authority weight of web document x and y_p is the hub weight. E isthe set of links (edges). Iteratively update the authority and hub weights of every web document, using Eq. (1) and (2), and sort the web documents in decreasing order according to their authority and hub weights, respectively, we can obtain the authorities and hubs of the topic.

III. EXISTING SYSTEM

Normally, web search engine receives query from the user and returns a list of web documents to them. The web search results may be displayed based on the content similarity, relevancy of keywords, hyperlink structure and web server logs. Conventional search engines provide users a list of non-classified web documents based on its ranking algorithm. However, sometimes these search results are far from user's satisfaction.

To provide more relevant web document to users to satisfy their need an Intelligent Cluster Search Engine (ICSE) [8] was developed. This system provided to the user a set of taxonomic web pages in response to a user's query and filters out the irrelevant pages. The following fig.3 shows the process of ICSE.

In this system, user's query is given to the meta-search engine. Then the clustered document set is created based on the given knowledge base and the clustering algorithm of ICSE. CA-ICSE [8] algorithm is used to cluster the web pages, which increases the relevancy of search results and reduces the computation time. This algorithm can be executed in two steps such as: compute the similarity and cluster the pages based on similarity. ICSE system consists of four modules such as: meta-search engine, meta-directory tree, web pages clustering, topic generation [8].

- ***Meta- search engine***

This module uses information extraction technology to parse the web pages and analyze the HTML tags. Stemmer is used to discard the common morphological and inflectional endings and Stop word to discard worthless words, and then the web pages will be converted to a unified format.

- ***Meta- directory tree***

In order to cluster the returned web pages rapidly, propose a novel clustering algorithm which uses meta-directory tree as the knowledge base for reducing the computation time required for clustering and enhancing the quality of clustering results.

- ***Web pages clustering***

Traditional clustering and classification technologies classify data without a knowledge base. It takes a lot of computation time to find classified results. To avoid this problem, it uses directory-tree approach which can not only cluster the web pages quickly but also assign a meaningful label to each group of classified results.

- ***Topic generation***

This module assumes that the words in the web page at the beginning and at the end parts are more important than in the middle part.

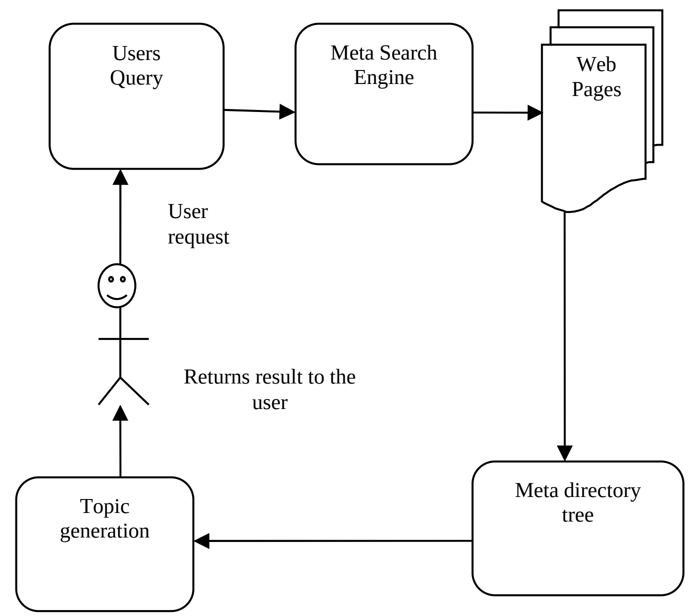


Figure 3: Design of Intelligent Cluster Search Engine

IV. PROPOSED SYTEM

In the proposed system, K-Mean clustering algorithm is used for information retrieval. K-Means clustering is more efficient in order to improve the relevancy rate of search results and also in saving computation time. The relevancy rate using CA-ICSE is decreased due to the similarity check between the documents using TF-IDF depending only on the contents. i.e. only the number of occurrences of a given word is compared in each document. So, in some documents the given word may have very low occurrence frequency and in other documents the word may have very high occurrence frequency. Based on ranking the documents are displayed in sequence which may have less similarity documents with highest priority and more similar documents may have least priority.

The least similar documents with high priority may lead to dissatisfaction of the user's needs. So the relevancy rate of documents must be increased in order to satisfy the needs of the users. The efficient way to improve the relevancy rate involves the use of K- Means Clustering algorithm [8].

In the proposed system by using K- Means Clustering algorithm, the Hub and Authority web documents are grouped based on the threshold given to the cluster and similarity measure. Based on the threshold value of each cluster the documents are selected and other are discarded. After clustering process, when a user requests for a query only the cluster with highest threshold is displayed to the user. This increases the relevancy rate and reduces the search space and

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

processing time. The following fig.4 shows the design of the proposed work. The proposed system works as follows:

- Enters a query onto the interface of search engine.
- Retrieve Hub and Authority documents for a Query.
- Decide the threshold value and compute the similarity of web document for relevancy by considering the weight of attributes in a data object.
- Once the weight is calculated, threshold value for clusters is assigned. According to the threshold values ,the documents clusters with most relevant, relevant and irrelevant clusters.The document which has weight with the centroid is assigned to the cluster and those doesn't support are discarded away from the cluster. The process is repeated until all the obtained results are clustered
- The IR system then receive only most relevant document to the user for a query

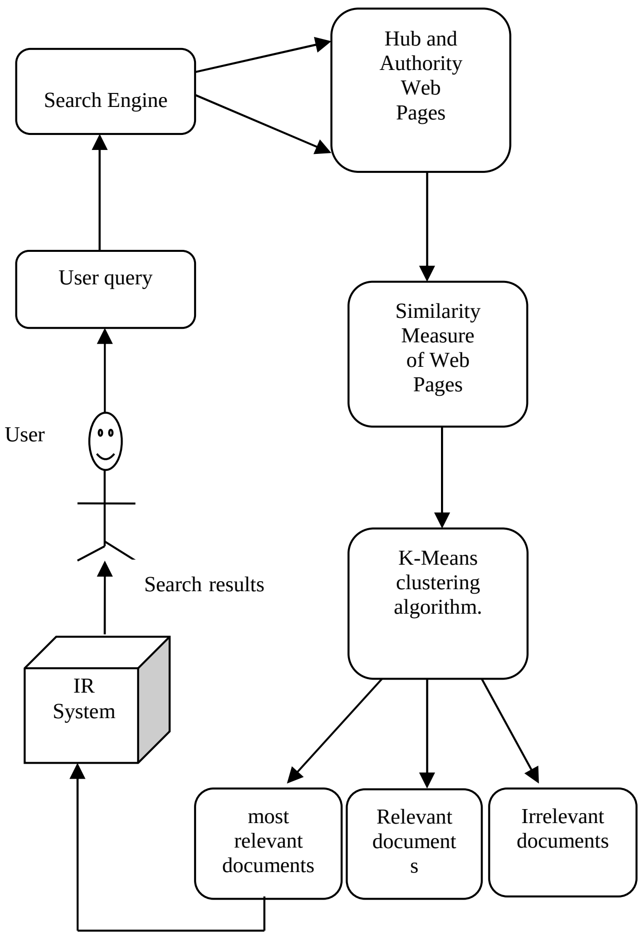


Figure.4 Design of proposed system

V. CONCLUSION

In this paper, an approach for clustering hub and authority web documents has been proposed. In which the similarity between the documents can be compared by considering the attribute properties of a data object (web document) not just by the contents of a document. All the documents are compared and the resultant clusters are formed by using K-Means clustering algorithm which improves the relevancy rate and processing time and search space significantly.

REFERENCES

[1] S. Brin and L. Page. "The anatomy of a large-scale hypertextual Web search engine", Computer Networks and ISDN Systems, 30(1–7):107–117, 1998.

[2] Preeti Chopra, Md. Ataullah, "A Survey on Improving the Efficiency of Different Web Structure Mining Algorithms" International Journal of Engineering and Advanced Technology (IJEAT), ISSN: 2249 – 8958, Volume-2, Issue-3, February 2013

[3] Laxmi Choudhary and Bhawani Shankar Burdak, "Role of Ranking Algorithms for Information Retrieva"l, International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.4, pages 203-220, July 2012

[4] Wenpu. Xing and Ali Ghorbani, "Weighted PageRank Algorithm", Proc. Of the Second Annual Conference on Communication Networks and Services Research, IEEE, 2004.

[5] Raymond Kosala, Hendrik Blockee, "Web Mining esearch: A Survey", ACM Sigkdd Explorations Newsletter, Volume 2,June 2000.

[6] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment" *Journal of the ACM*, 46(5):604–632, September 1999.

[7] Mr. Dushyant Rathod, "A Review On Web Mining", International Journal of Engineering Research and Technology (IJERT), Vol. 1, Issue 2, Pages 21-25, 2012.

[8] M.Sathya, J.Jayanthi, N. Basker, "Link Based K-Means Clustering Algorithm for Information Retrieval", International Conference on Recent Trends in Information Technology (ICRTIT), 978-1-4577-0590-8, IEEE MIT, Anna University, Chennai. June 3-5, 2011

[9] M. Steinbach, G. Karypis, and V. Kumar, "A Comparison of Document Clustering Techniques," Proc. KDD-2000 Workshop on Text Mining, Aug. 2000.

[10] Yitong Wang and Masaru Kitsuregawa, "Use Link-based Clustering to Improve Web Search Results", 0-7695-1393, IEEE, 2002.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

(IJCSIS) International Journal of Computer Science and Information Security,
Vol. 14, No. 3, March 2016

Dorsal hand vein identification

Sarah HACHEMI BENZIANE
Irit, University Paul Sabatier France
Simpa, University of Sciences and Technology of Oran
Mohamed Boudiaf Algérie

Abdelkader BENYETTOU
Simpa, University of Sciences and Technology of Oran

***Abstract*— In this paper, we present an competent approach for dorsal hand vein features extraction from near infrared images. The physiological features characterize the dorsal venous network of the hand. These networks are single to each individual and can be used as a biometric system for person identification/authentication. An active near infrared method is used for image acquisition. The dorsal hand vein biometric system developed has a main objective and specific targets; to get an electronic signature using a secure signature device. In this paper, we present our signature device with its different aims; respectively: The extraction of the dorsal veins from the images that were acquired through an infrared device. For each identification, we need the representation of the veins in the form of shape descriptors, which are invariant to translation, rotation and scaling; this extracted descriptor vector is the input of the matching step. The optimization decision system settings match the choice of threshold that allows to accept / reject a person, and selection of the most relevant descriptors, to minimize both FAR and FRR errors. The final decision for identification based descriptors selected by the PSO hybrid binary give a FAR =0% and FRR=0% as results.**

Keywords- Biometrics, identification, hand vein, OTSU, anisotropic diffusion filter, top & bottom hat transform, BPSO,

I. Introduction

THIS last years, the research community show an increasing interest to biometrics. Although, the biometrics commercial products was born during this decade which is driven by security issues.

Biometrics has several modalities, they are classified in three principal techniques: biological technique as AND[1]; physiological technique as fingerprint[2], face[3], iris[4] and hand[5]. And finaly the behavioural technique: as keystroke [6], gait [7]. Most of the works was made in the visible spectrum. In this last years, some research start interesting in the non-visible spectrum like infrared images used in the hand vein identification/authentication.

Infrared thermography (IRT), thermal imaging, and thermal video are examples of infrared imaging science. That imagery produced as a result of sensing electromagnetic radiations emitted or reflected from a given target surface in the infrared position of the electromagnetic spectrum (approximately 0.72 to 1,000 microns). Many researchers used the thermal imaging for face and hand authentication, they got efficient results. Although the infrared imaging is less expansive and present too efficient results for that it was used in the both hand [9] and face [8] recognition.

In this paper, we present a work about the use of active near infrared imagery for the feature extraction of dorsal hand vein. This features shaped the dorsal venous network of the hand. This last is used for person identification/authentication. Many works was made for this feature extraction. Especially, [10] who used single triangulation of hand vein images and simultaneous extraction of knuckle shape information. In [11], the palm and dorsal veins are considered as texture samples being automatically extracted from the user's hand image. A 2D Gabor filter is employed for texture feature extraction. When [12], present the enhancement's step of the SAB11 Data Base for adaptive feature extraction method of the dorsal hand vein biometrics; which is the discrete wavelet transform.

The biometrics word has a larger meaning in the study of identification/authentication persons from a number of characteristics. It is a Mathematical analysis of biological and/or behavioral characteristic of a person to determine his identity decisively. Biometric modalities are based on principle characteristics recognition as Fingerprint [26], face [41], iris[31], retina[42], hand, keystroke, voice and vein; they provide irrefutable proof of the identity of a person by their biological uniqueness characteristics distinguishing one person from another. The hand vein biometrics has emerged as a promising component of the biometric study [23], [12], [39], [5]. Each Biometric system has a processing chain has carried particular the hand vein systems to get the final decision [15]. The biometric vein pattern presents a very high level of security, to date, no way to defraud, called biometrics "contactless". The rest of the paper is organized as follows: First, we give an overview of prior researches relevant to Hand vein biometric. In section 2, we made a description of the proposed system; Section 3 presents enhancement of the quality of the database used for better vein feature extraction which is detailed in Section 4. Section 5; show how to get the vector feature extraction. Section 6, gives a brief description about the hybridation of the BPSO. Section7 presents the experimentation and results of the proposed system; conclusions are drawn in Section 8.

II. Prior Work

In visible light, the veins are not apparent. Indeed, a multitude of other factors, including the surface characteristics

 https://sites.google.com/site/ijcsis/
ISSN 1947-5500

such as moles, warts, scars, pigmentation and hair can also hide the image [29]. Fortunately, the use of the infrared light eliminates most unwanted surface features[35]. Required parameters to obtain good quality data are listed below [1][29]:

• The light affects the quality of the image obtained with the exception of no IR filter.

• The temperature of the ambient environment must be neither too hot nor too cold, around the human body temperature.

• The distance between the sensor and the object should be sufficient for a good acquisition.

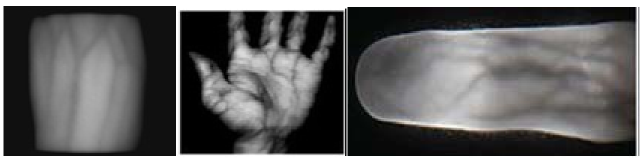


Fig. 1 Dorsal, Palmar and Fingerprint veins

Starting with Jackson W. Wegelin patent [11], includes a dispenser controller coupled to a memory unit, which includes a database of previously-stored vein patterns. A vein-pattern sensor maintained by the dispenser images the unique vein pattern of a user’s hand without contact. A recent study proposed using a three dimensional biometric scanner (1) for the capillary mapping of the palm of the hand (2) incorporates two image sensors; configured for obtaining a stereoscopic image of a vascular map and where for each image corresponding to each wave length, the depth of each point on the plane is known [2]. Some multimodal biometric systems capture plamprint/ finger vein [20][7], hand geometry/vein [4][40]. Other systems capture the palmar veins [8]. [24] Based on registering finger vein information[13], and it is discriminated of which finger the finger vein information is to be registered, on the basis of the photographed image. There is too central combining several modalities as [30] which comprise a central command station in signal communications with a series of blasting machines. Command station has a biometric analyser unit and an authorizing means. The blasting apparatuses have enhanced security features by including biometric analysis of specific biological features of an authorized blast operator to generate a known biometric signature. The biometric signature can be derived from a fingerprint scan, a recognition scan of a hand, a foot, an iris or a retina, a skin spectroscopy analysis, a finger vein pattern analysis, a voice recognition analysis, or a DNA fingerprint analysis. In order to get areas, improve the quality of the image, extract veins from hand, we need some techniques:

a) Format conversion JPEG BMP: the conversation JPEG BMP is required. Indeed, the main advantage of BMP image quality is provided as BMP format is not compressed and therefore no loss of quality. Against by the JPEG format is compressed and therefore quality lost [25].

b) Enhancement: The resulting image may not contain noise as tasks, blobs, dust ... ect. Different filters can be applied to eliminate the noise and enhance the image, but if the pictures have a good quality, this step is not required [34]. In [36], the clearness of the vein pattern in the extracted ROI varies from image to image; they use a 5x5 Median Filter to remove the speckling noise in the images and a 2-D Wiener filter to the ROI image to suppress the effect of high frequency noise. [27] uses a various contrast enhancement techniques in order to compare which gives the best results, the study is very interesting.

c) Converting the color image into a gray level: Converting a color image into a grayscale means that the image size will be reduced from 24 bits per pixel (color image) to 8 bits per pixel (grayscale image) [14] [16]. Instead of having three matrices that represent the level of colors (red, green, blue) for each pixel, we have just a single matrix that represents the gray level for each pixel, which reduces the processing time [37] [38].

d) Binarization: Binarization is the segmenting the image into two levels; object (hand region) and background; most of the time the object segment which is the region of interest (ROI) in white and the background segment in black [3] [33] [17]. After the binarization, there is the most difficult step which is the feature extraction. Some researchers add a step in this module [17] [16] [25] [28]. Some works use the minutiae features extracted from the vein patterns for recognition, which include bifurcation points, ending points and the position and orientation of minutiae points [11] [10]. [21][22] Uses it with the vein finger, when [36] uses it with the dorsal hand vein. In [33] the feature extraction was based on the geometry veins. The figure below shows these minutiae (playback direction: from left to right).

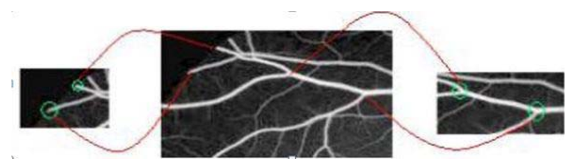


Fig. 2 End Points, veins branching points[21]

Identification/Authetication phase or then verification is only classification [46] of items into two classes. The image of the veins that was extracted in the previous phase allow us to create a database of prototypes with (s) models that are in the base (template) by Authenticating the identity of an individual, will either accept the person, or reject it. Instead of the identification, the system will identify the right person. In order to evaluate their system testing performance, [3] uses a dataset of 500 persons of different ages above 16 and of different gender, each has 10 images per person was acquired at different intervals, 5 images for left hand and 5 images for right hand. [43] used correlation and template matching as a recognition algorithm whether [44] used Using Principle Component Analysis (PCA). All the works are resumed in the table1.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Table 1 Survey of hand vein biometrics

III. PROPOSED SYSTEM DESCRIPTION

This paper deals with a new biometric Identification approach. The main contributions of this paper can be summerized as follow:

- Representation of the veins vector in the form of shape descriptors,they are invariant to translation, rotation and scaling. The classification is done based on these descriptors.
- optimization decision system settings match the choice of threshold that allows to accept / reject a person, and selection of the most relevant descriptors, to minimize both FAR and FRR errors.
- The integration of hybrid binary PSO for solving the bi-objective optimization problem (FAR and FRR minimized). This meta-heuristic decision provides greater credibility by introducing the notion of subjective parameters (formulated by the decision maker), corresponding to the weight assigned to the FAR and FRR.
- Identification based descriptors selected by the PSO hybrid binary.

From The dorsal vein hand image are extracted contours used for the image normalization and segmentation of region of interest (ROI) which is detailed in Sections 2. The extraction of hand vein vector from ROI images is described in Section 3. The extraction feature from the hand vain vector and the identification are detailed in Sections 4 and 5 respectively. The experiments and results of this work are presented in Section 5 which is followed by the discussion in Section 6 and the main conclusions of this paper.The main architecture of our identification biometric system consists of three modules: Enhancement module, feature extraction and classification.The first two modules were implemented on CPU.

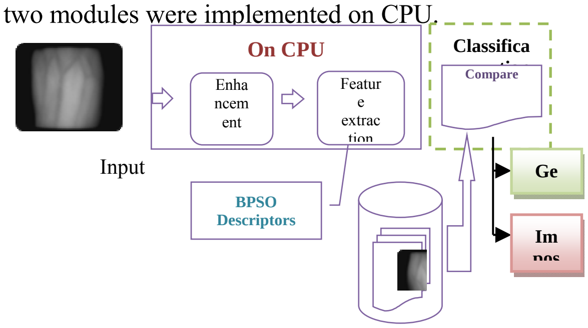


Fig.3 Block diagram of our decision support system

The functional architecture of the decision system matches the biometric identification with dorsal hand vein developed; is shown below.

IV. IMAGE QUALITY AMELIORATION

The NCUT databases images are images taken of the hands at a distance. Thus, the image is composed of two parts by hand and background surrounding To avoid wasting time calculation by processing a non-interesting area which is the background of the image. We applied a binarization to extract the area that interests us (hand). The binarization step allows us to divide the

Reference	Dimension	Thresholding	Binarization	Vein	Minutiae's	Classification	Number	Performance
				Extraction	extraction		Dataset	
[9]	160x120	Gaussian low	Local	Local	No	/	10.000	FAR 0.01%
		Pass and high pass	Thresholding	Thresholding				
[18]	Combinaison	Median	Local	Wavelet	No	/	30	FRR 1.5%
	Multi resolution		Thresholding	Transform				FAR 3.5%
[32]	/	Median	Local	Local	No	Hausdorff	12	FRR 0%
		Gaussian	Thresholding	Thresholding		Distance		FAR 3.0%
[6]	/	GSZ	No	Gabor	Cross	KNN with	/	/
		Shock		Thresholding	number	Euclidean		
[19]	640x480	Median	Threshold=0	Skeletonisation	/	/	/	/
	Special Median							
[19]	640x320	Guassian	SIFT	SIFT	/	Euclidian	24	EER=0%
		LOW Pass				Distance		

image in two areas: the hand area(white) and the background (black). We applied the algorithm of OTSU; see algorithm below:

Upload image;
Convert from color to grayscale;
Calculate the histogram of the image;
Normalization of the histogram;
Mean (0) =0;
Var (0) =0;
If k<= 255
  Update mean (k);
  Update car (k);
  Calculate $S^2(k)$;
Else
  Threshold Binarization, T=k
  $S^2(k)=max(S^2(k))$;
  If $I(x,y)>T$
  $I(x,y)>T=1$
  Else
  $I(x,y)>T=0$

Fig. 4 Algorithm of OTSU for the binarization

Since the background is useless, we must eliminate it and keeping only the white pixels in the image that is to say; the area of the hand; see figure 3.

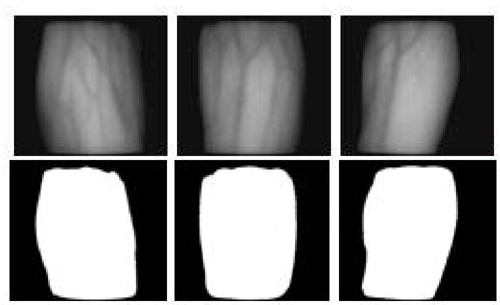


Fig.5 binarization for hand area extraction

In order to extract the area of interest (ROI) that contains the veins, we calculated the gravity hand center; as shown in equation 1 and 2. this operation aims to eliminate the bottom (the image size reduction) and have more accurate results.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

$$X_g = \frac{\sum_{i,j} i*I(i,j)}{\sum_{i,j} I(i,j)}; \quad (1)$$

$$Y_g = \frac{\sum_{i,j} j*I(i,j)}{\sum_{i,j} I(i,j)} \quad (2)$$

As the images of the veins were not clearly distinguished; we improved the contrast by a linear transformation. The results of different people are shown in figure 6.

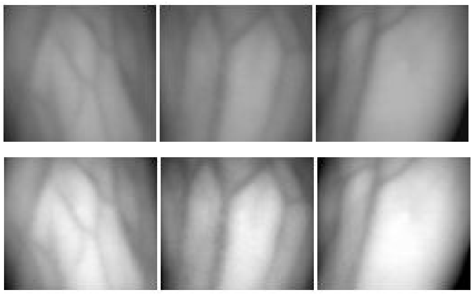


Fig. 7 Extraction and enhancement

V. DORSAL HAND VEIN FEATURES

the most important features that can be extracted from the back of the hand are so called dorsal venous network. The dorsal venous network of the hand is a network of veins formed by the dorsal metacarpal veins (Figure 1.[13]). In anatomy, the ulnar veins are venae comitantes for the ulnar artery. They mostly drain the medial aspect of the forearm. They arise in the hand and terminate when. Dorsal venous architecture In 82% of 300 individuals a large vein passed proximally from the center of the concavity of the dorsal venous arch to terminate in 65% in the cephalic vein, and in the remaining 17% in the basilic vein. These venous network are unique to each individual for that it is used as a biometric technique for person identification/authentication.

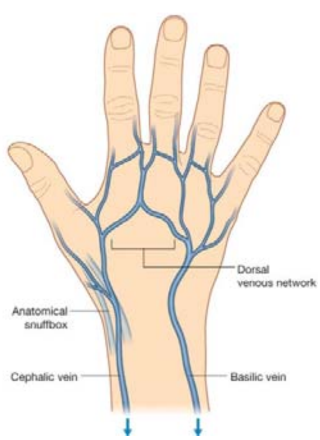


Figure 1: Dorsal Hand Veins | Varicose Veins [13]

The absorption of the skin to certain wavelengths in the near infrared spectrum allows us to extract these features. Many works resumed the research made in this field. The work presented in this paper is based on the SAB'11 and SAB'13 database. This is database acquired from a biometric infrared device representing hundreds of dorsal hand veins samples; for male and female of different ages; sensitive to the infrared 850-900 nm waves [14]. This spectrum is indented for the current feature extraction because of the blood oxy-hemoglobin is higher than deoxy-hemoglobin and skin water. Which allow to the near infrared light to penetrate the skin and to be absorbed by the blood of the veins Figure 2.

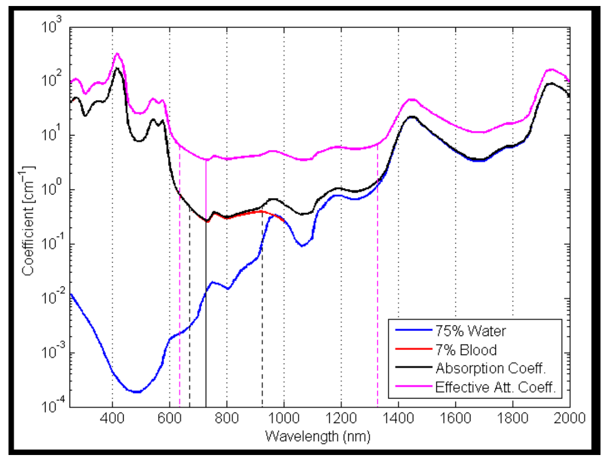


Figure 2 : Spectra for veins (SvO2 ≈ 60%). Absorption coefficient: λmin = 730 nm; NIR window = (664 - 932) nm. Effective attenuation coefficient: λmin = 730 nm; NIR window = (630 - 1328) nm.

In this part, we show two distinguish results based on two distinguish techniques:

1. First Feature extraction results :

Once we extract ROI, we proceeded to a binarizaion thresholding operation to divide the image into two levels: black background and white veins, by the method of integral image. Based on the following algorithm:

Calculating the integral of all images (with H the height and L the width);
Compute the new center (x, y) of the window;
if x<H
Calculating the local sum of the integral;
Calculating the local mean;
Calculates the standard deviation;
Calcultes the threshold T for the center of the window;
If $I(x,y)>T$
$I(x,y)>T=1$
Else
$I(x,y)>T=0$

Fig.8 The integral image for binarization

The problem with this method was how to choose the size of the window (w * w) and coefficient k .It's why we applied several tests in the size of the window, setting k = 0 on several people .The following figure shows the results for the same person:

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Fig. 9 Binarization by integral image for the extraction of dorsal hand veins

From the results, we found that the window size with (55 * 55) gave us the best visualization of the veins, so we took the window size (55 * 55) and coefficient k = 0 for binarization. To improve the quality of the binary image, we applied the morphological dilation operation, which will allow us to remove black areas of the image. The results are shown in the following figures:

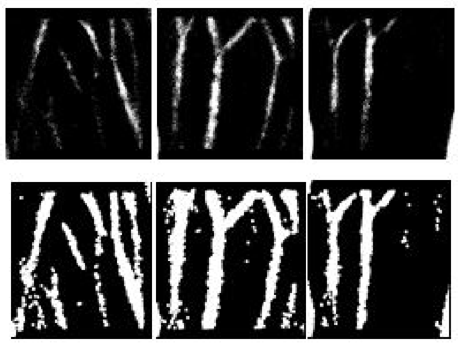


Fig. 10 Before and After Dilation with (7x7)

According to the results, we noticed that there still have isolated pixels, so it must be eliminated by applying a filter. We chose the Median filter because usually it is the response filter. The results are shown in the figure 10.

Fig. 11 Vein Binary Image after Dilatation

Fig.12 Vein Binary Image after Dilatation Window (7*7)

VI. VECTOR FEATURE EXTRACTION

This step is intended to represent the veins by color descriptors, texture, shape, or the combination of both these descriptors. since the comparison between the images is done by these descriptors. For this, we chose the shape descriptors, cause they are invariant to rotation, translation and scaling. So we opted for the method of Hu moment invariants, which will extract seven descriptors of shapes, from the binary images of dorsal veins of the hand. The corresponding organigram is presented figure 10:

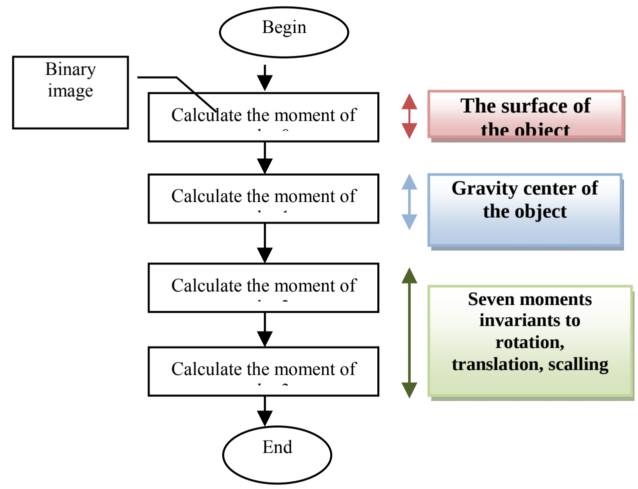


Fig.13 feature extraction by the method of moment invariants HU

2. Features vein extraction based anisotropic diffusion filter

To get the extraction features characterizing the venous network, we suggest the following steps that are summarize in the block diagram of Figure 4:

1. Improve contrast enhancement; using the anisotropic diffusion [15][16][17].
2. Get the feature extraction by the hat morphological filtering [18][19].

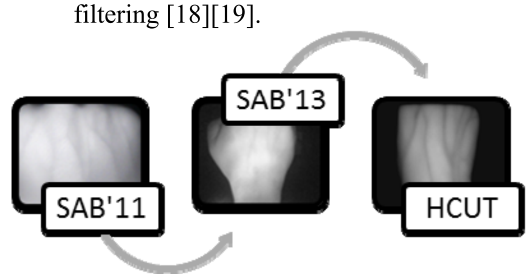


Figure 3: Samples of the infrared dorsal hand vein images from SAB'13.

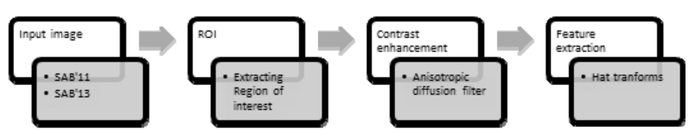


Figure 4:Block diagram of proposed system

VII. CONTRAST ENHANCEMENT

In physiological hand biometrics, the quality of an image is determined by two criteria, namely hand (background) and veins (stripes). Backgound result from isotropic inhomogeneities of the density distribution, whereas stripes are an anisotropic phenomenon caused by adjacent veins pointing in the same direction. Anisotropic diffusion filters are capable of visualizing both quality-relevant features instantaneously. For getting an appropriate parameter choice, we can achieve isotropic smoothing at clouds and diffuse in an anisotropic way along veins in order to enhance them.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

However, if one wants to visualize both features separately, one can use a fast pyramid algorithm based on linear diffusion filtering for the background, whereas stripes can be enhanced by a special nonlinear diffusion filter which is designed for closing interrupted lines.

Perona and Malik propose a nonlinear diffusion method for avoiding the blurring and localization problems of linear diffusion filtering [20][21]. They apply an inhomogeneous process that reduces the diffusivity at those locations which have a larger likelihood to be edges. This likelihood is measured by $|\nabla u|^2$. The Perona-Malik filter is based on the equation:

$$\partial_t u = div(g(|\nabla u|^2)\nabla u) \quad (1)$$

And it uses diffusities such as:

$$g(s^2) = \frac{1}{1+s^2/\lambda^2} \quad (\lambda > 0) \quad (2)$$

The proposed diffusion process encourages intraregion smoothing. The mathematical framework for anisotropic diffusion is given by the equation below:

$$\frac{\partial}{\partial t} I(\bar{x},t) = \nabla\cdot(c(\bar{x},t)\nabla I(\bar{x},t)) \quad (3)$$

Where:

$I(\bar{x},t)$: image;

$\bar{x}$: image axes (i.e. (x,y));

t : iteration step;

$c(\bar{x},t)$: diffusion function; [20] proposed two functions Equation (4) and (5):

$$c_1(\bar{x},t) = \exp\left(-\left(\frac{|\nabla I(\bar{x},t)|}{k}\right)^2\right) \quad (4)$$

$$c_2(\bar{x},t) = \frac{1}{1+\left(\frac{|\nabla I(\bar{x},t)|}{k}\right)^{1+\alpha}} \quad | \alpha > 0 \quad (5)$$

k is the diffusion constant.

Feature extraction describes the relevant shape information contained in a pattern so that the task of classifying the pattern is made easily by a formal procedure. In pattern recognition and in image processing, feature extraction is a special form of dimensionality reduction. The main goal of feature extraction is to obtain the most relevant information from the original data and represent that information in a lower dimensionality space. Features represent important components of the venous network from the enhanced image. Knowing that only the central area of the image is interesting, we extract a ROI for the feature extraction input. After what, we apply mathematical morphology to extract the veins network.

Top-hat transform is an operation that extracts small elements and details from given images. There exist two types of top-hat transform: The white top-hat transform is defined as the difference between the input image and its opening by some structuring element; the black top-hat transform (sometimes called the *bottom-hat* transform) is defined dually as the difference between the closing and the input image. Top-hat transforms are used for various image processing tasks, such as feature extraction, background equalization, image enhancement, and others.

First, we apply top and bottom hat transforms to extract the desired features basing on a suitable structuring element B that is bigger than the width of the subcutaneous vessels in the image.

The algorithm is based on combining image subtraction with openings and closings results in **top-hat** and **bottom-hat** transformations. The ***top-hat*** transformation of a gray-scale image *f* is defined as *f* minus its opening:

$$T_{hat}(f) = f - (f \circ b) \quad (6)$$

Similarly, the ***bottom-hat*** transformation of a gray-scale image *f* is defined as the closing of *f* minus *f*:

$$B_{hat}(f) = (f \bullet b) - f \quad (7)$$

Then we substract the two obtained image. The structure element B in our experiments in a disk diameter 4

The top and the bottom hat transforms are given below:

$$I_f = B_{hat}(I,B) - T_{hat}(I,B) \quad (8)$$

Our experiments were lead on the SAB'13 [14] database with a All in one HP machine; that's configuration is Intel (R) Core (TM) i3-3240 CPU @3.40GHz 3.40GHz. The SAB'13 database was conducted with a built biometric dorsal hand vein device which is based on a camera that has a good sensitivity in the near infrared spectrum. A lighting system with hundreds infrared Led's emitting in the spectrum 850nm were used. In this paper, we proposed techniques which allow us to get the extraction of wanted vein network features.

The previous works shows that the uniqueness of the network vein let us use this modality as way to get person identification/authentication. Knowing that for each context and image, the central area of the image is the most interest part to use.

In our tests we start first across all the used databases with different variations in the diffusion constant and number of iterations until finding the efficient result, with below parameter. The results are shown in Figure 5.

IM - gray scale image (MxN).

Num_iter - number of iterations.

Delta_t - integration constant (0 <= delta_t <= 1/7).

Usually, due to numerical stability this parameter is set to its maximum value.

Kappa - gradient modulus threshold that controls the conduction.

Option - conduction coefficient functions proposed by Perona & Malik:

1 - c(x,y,t) = exp(-(nablaI/kappa).^2), privileges high-contrast edges over low-contrast ones according to Equation 4.

2 - c(x,y,t) = 1./(1 + (nablaI/kappa).^2), privileges wide regions over smaller ones according to Equation 5.

Where:

num_iter = 180;
delta_t = 1/7;
kappa = 100;
option = 2;

Original image	Anisotropic diffusion filter	Features extracted (Vein line network)

SAB'11 / SAB'11 / SAB'13 / SAB'13 / NCUT / NCUT

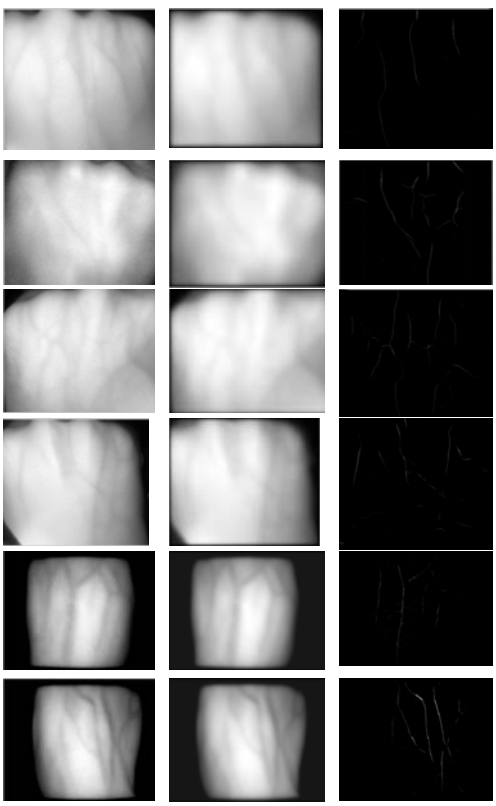


Figure 5: Dorsal hand vein results for three databases.

The Figure 5, show more the efficiency of the diffusion filter used for getting rapidly and efficiently the dorsal venous network. Nest figure 6, permit to see more the results.

SAB'11

SAB'13

NCUT

Figure 6: Anisotropic diffusion filtered image

VIII. HYBRIDIZATION OF BPSO WITH MOMENT INVARIANTS HU

The disadvantage of the seven Hu moment invariants is that they are sensitive to noise, in addition they are very hungry time calculated. So in order to select only the moments that minimize both FAR and FRR errors, we will make a hybridization of BPSO (binary PSO) with the method of invariant moments of HU. This hybridization has two main goals:

- Select less time instead of seven times like shape descriptors dorsal hand veins, which minimize both errors FAR and FRR.
- To have an optimal threshold that can accept or reject people.

This hybridization was never done before. It is presented as follows:

Step 1: Initialize settings BPSO (N: the number of particles the number of iterations).

Step 2: Initialize the positions X and speed V.

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{iM} \\ x_{21} & x_{22} & \cdots & x_{iM} \\ . & . & . & . \\ . & . & . & . \\ x_{N1} & x_{N2} & \cdots & x_{iM} \end{pmatrix} \quad (3)$$

$$x_{im} = randint()$$

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

$$V = \begin{pmatrix} v_{11} & v_{12} & \dots & v_{iM} \\ v_{21} & v_{22} & \dots & v_{iM} \\ . & . & . & . \\ . & . & . & . \\ v_{N1} & v_{N2} & \dots & v_{iM} \end{pmatrix} \quad (4)$$

$$v_{im} = -vmax + 2vmax * rand() \quad (5)$$

where :

i is the particle index

m the m^{th} dimension

x_{im} The m^{th} selected time of particle i

$randint()=$ { means that the moment has been selected 0 else.

M The number of selected moment; in our case M=7

N The number of particules

$vmax$ The number of changement; in our case = 7 invariant moments

$rand()\varepsilon[0,1]$

Step 3 : The comparison of the images are made on the basis of moments that were selected by BPSO according to the chart below, every person in our case is a class.

Step 4 :

Upload Input class(binary vein image) Extraction of the seven invariant moment of Hu Selestion of the moment by BPSO Calculate $d1$: Distance between the input class and all DBB classes Find the minimal Class C Separate class C into two groups, and calculating the distance $d2$ between them by the $Dist$ **if** $\frac{d1}{d2} \leq T$

Fig. 14 Algorithm of the classification by the hybridation

Where the distance between the images of each person and different people (classes) is calculated by the following equation:

$$Distance = \frac{1}{n_1 * n_2} \sum_{i=1}^{n1} \sum_{j=1}^{n2} dist(A_i; B_j) \quad (6)$$

Where n_1 and n_2 are the number of an image in class A and B resp., $dist(A_i; B_j)$ is the Euclidean distance between the image A_i and the class B.

Step 5 : Update the fitness function of each particle, which has two objectives:

- Minimize FAR error: The false acceptance rate.
- Minimize FRR error: The false rejection rates.
- The corresponding function is

$$MinimizeE = CFA * GFAR + CFR * GFRR \quad (7)$$

$$CFR = 2 - CFA \quad (8)$$

Where the parameters of the function fitness are defined as:

GFAR is FAR
GFRR is FRR
CFA is the cost of false acceptance
CFR is the cost of false reject
EER is the error rate

Both the FAR and FRR objectives are defined by the following equation:

$$FRR = \frac{Numberofrejectedgenuine}{Totalgenuinewhoaccessed} * 100 \quad (9)$$

$$FAR = \frac{Numberofimpostorsaccepted}{Totalimpostorswhoaccessed} * 100 \quad (10)$$

Step 6: If the fitness function of each particle is better than the previous best fitness function, then the current position is the best previous position and the current fitness function is its previous best fitness function.

Step 7: Assign the minimum fitness function all the best fitness functions of each particle, the overall fitness function. The positions of this function will be assigned to the best overall position. Step 8: Update the speed. Step 9: Update position. Step 10: Repeat steps 3, 4, 5, 6, 7, 8 until stopping criterion.

IX. EXPERIMENTATION AND RESULTS

The information about the database that we used were acquired from the NCUT Database[45]. The information about the database used are summarized in the Table 2:

Parameters	Definition
Number of person in the DBB	102 (52 women; 50 men)
Number of image per person	10 for the right and for the left hand
Number of image in the DBB	2040
Number of person (class) for the training	50(6 images for each person) and the remaining for the test.

Table 2 Database Information

The parameters used in this experiment for BPSO are summarized in the table 3, according to :

Number of iterations	50
Number of particle	10
$c1$	0.9
$c2$	1
w_{max}	1.9
w_{min}	0.4
r_1	0.5
r_2	0.5

Table 3 Parameters used

For the fitness function, we varied the cost of CFA in the range [0.1,1.9] and the threshold that allows to know if it is a genuine or fake. The following figures show the error evaluation FAR, FRR and the objective function (error rate) during the variation of the threshold.

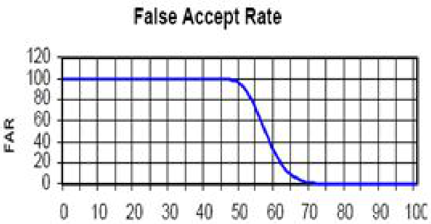


Fig. 13 False Accept Rate FAR

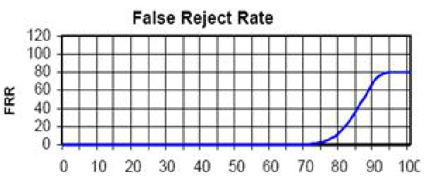


Fig. 14 False Reject Rate FRR

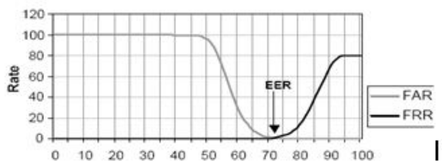


Fig. 15 Equal Error Rate EER

Through the histogram in Figures above, the two errors FAR and FRR were equal with the threshold value is 72%. So the threshold that gave us the best results was 72%, it what we considered it.

The results obtained by the hybridization of BPSO with Hu moment invariants for the selection of the best times that minimize both FAR and FRR errors are listed in the following table.

CFA	$\theta 1$	$\theta 2$	$\theta 3$	$\theta 4$	$\theta 5$	$\theta 6$	$\theta 7$	FAR	FRR	EER
.1	1	0	0	1	0	1	0	0	0	0
.3	0	1	1	1	0	0	1	2	0	0.6
.4	1	1	1	1	0	0	0	0	0	0
.6	0	1	1	1	1	1	0	0	0	0
.9	0	1	1	1	1	0	0	0	0	0
	1	1	1	1	0	0	1	0	0	0
.6	1	1	1	1	1	0	1	0	0	0
.9	1	1	0	0	0	0	0	0	0	0
Mean	0.73	0.73	0.68	0.69	0.57	0.36	0.57	0	0	0

Table 4 : Results of Hu moment invariants selected by BPSO

From the table above, the results obtained by hybridization of BPSO and Hu moment invariants, were able to reach a figure of 0% for both FAR and FRR errors with less time selected instead of seven and a threshold of 72% by varying the CFAR and CFRR .Another remark cost, we observed from the table above, is that the moments $\theta 1; \theta 2$ were the most selected by BPSO times, so for confirm that, we did a manual selection times with or without these two moments $(\theta 1; \theta 2)$.

The results are shown in the following table:

Manual Selection of the moments							FAR %	FRR %	Min
Φ_1	Φ_2	Φ_3	Φ_4	Φ_5	Φ_6	Φ_7			
1	1	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	2	0	3.78
0	0	1	1	1	1	1	2	1	3.79
1	0	1	1	1	1	1	2	2	4
0	1	1	1	1	1	1	0	2	0.2

Table 5 : Results of the manual selection

From the table, we see that the absence of one of these two moments or both, or their presence with both other times the error rate increases. However, the presence of only these two times both the error rate becomes 0%. Therefore, we have taken only two times $(\theta 1; \theta 2)$ to represent the veins by two shape descriptors.

X. Conclusion

In this work we propose a new technique for extracting dorsal hand vein as physiological features in the aim of biometric recognition. In this work, we used the built database of near infrared dorsal hand with the corresponding features; known more as SAB'11, SAB'13 and NCUT Benchmark. These features represent the subcutaneous dorsal venous network of the hand, although the superficial veins forming the median antebrachial vein of the dorsal hand. With the proposed technique, we get efficient results for our tests for the extraction of the required features. Those features are important in their use in biometric applications for getting person identification/authentication.

According to both measurement error rates FAR and FRR, we observe that we have had the best results with an error rate FAR and FRR = 0% among the work done. According to the execution time, we cannot compare our results to 100% because it depends on the type of used machine. In addition a few studies which have mentioned the execution time of the pretreatment phase, feature extraction and classification. Indeed, in biometrics, the authors focus primarily on minimizing both error rate FAR and FRR, after the execution time for biometrics is a type of soft real-time, that is to say, even if there's a delay in obtaining the result will not cause damage. However, it is clear to point out some very important aspects we have deduced by this project: It is not necessary to make the dorsal veins as skeletons to extract minutiae, because: to. This makes it very heavy preprocessing. In addition, once again it is difficult to extract the minutiae of the dorsal hand veins, because of the quality of the infrared camera used to acquire images of veins. b. May decrease the performance of FAR perspective FRR system, because if one of the minutiae algorithms does not detect it, we will file it in another category.

References

[1] Lamine HANINI Lakhdar LIMAM. Mise au point d'un système d'acquisition biométrique : SAB-11 Veines de la

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

main. Institute of Maintenance and Industrial Safety; University of Oran, 2011.

[2] D.N. Antequera, Delrios Sanchez, 3d biometric scanner, February 1 2012. EP Patent App. EP20,090,842,118.

[3] Ahmed M Badawi. Hand vein biometric verification prototype: A testing performance and patterns similarity. IPCV, 14:3–9, 2006.

[4] B.W. Beenau, D.S. Bonalle, S.W. Fields, W.J. Gray, C. Larkin, J.L. Montgomery, and P.D. Saunders. Hand geometry biometrics on a payment device, October 16 2012. US Patent 8,289,136.

[5] Sarah Benziane And Abdelkader Benyettou. Biometric technology based on hand vein. Oriental Journal of Computer Science and Technology, 2013.

[6] M Deepamalar and M Madheswaran. An improved multimodal palm vein recognition system using shape and texture features. International Journal of Computer Theory and Engineering, 2(3):436–444, 2010.

[7] Youssef Elmir, Zakaria Elberrichi, and Réda Adjoudj. Multimodal biometric using a hierarchical fusion of a person's face, voice, and online signature. Journal of Information Processing Systems, 10(4), 2014.

[8] S. Hama, T. Aoki, and M. Fukuda. Biometric authentication device and biometric authentication method, October 5 2011. EP Patent App. EP20,080,878,913.

[9] Sang-Kyun Im, Hyung-Man Park, Young-Woo Kim, Sang-Chan Han, Soo-Won Kim, Chul-Hee Kang, and Chang-Kyung Chung. An biometric identification system by extracting hand vein patterns. JOURNAL-KOREAN PHYSICAL SOCIETY, 38(3):268–272, 2001.

[10] Anil K Jain, Lin Hong, Sharath Pankanti, and Ruud Bolle. An identity-authentication system using fingerprints. Proceedings of the IEEE, 85(9):1365–1388, 1997.

[11] Anil Jain, Arun Ross, and Salil Prabhakar. Fingerprint matching using minutiae and texture features. In Image Processing, 2001. Proceedings. 2001 International Conference on, volume 3, pages 282–285. IEEE, 2001.

[12] Anil K Jain, Patrick Flynn, and Arun A Ross. Handbook of biometrics. Springer Science & Business Media, 2007.

[13] Fathima, A. Annis, et al. "Fusion framework for multimodal biometric person authentication system." IA ENG Int. J. Comput. Sci 41.1 (2014).

[14] R Kavitha and Little Flower. Localization of palm dorsal vein pattern using image processing for automated intra-venous drug needle insertion. International Journal of Engineering Science & Technology, 3(6), 2011.

[15] Adams Kong, David Zhang, and Mohamed Kamel. A survey of palmprint recognition. Pattern Recognition, 42(7):1408–1418, 2009.

[16] LOHYEE KOON. Implementation of hand vein biometric authentication on fpga-based embedded system.

[17] Ajay Kumar and K Venkata Prathyusha. Personal authentication using hand vein triangulation and knuckle shape. Image Processing, IEEE Transactions on, 18(9):2127–2136, 2009.

[18] Chih-Lung Lin and Kuo-Chin Fan. Biometric verification using thermal images of palm-dorsa vein patterns. Circuits and Systems for Video Technology, IEEE Transactions on, 14(2):199–213, 2004.

[19] Dana Lodrova, Radim Dvovrak, Martin Drahansky, and Filip Orsag. A new approach for veins detection. In Bio-Science and Bio-Technology, pages 76–80. Springer, 2009.

[20] E. Makimoto, K. Nomura, Y. Mizuno, N. Takamura, and H. Ogata. Biometric authentication apparatus and biometric authentication method, April 20 2011. EP Patent App. EP20,100,187,628.

[21] Suleyman Malki and Lambert Spaanenburg. Hand veins feature extraction using dt-cnns. In Microtechnologies for the New Millennium, pages 65900N–65900N. International Society for Optics and Photonics, 2007.

[22] Naoto Miura, Akio Nagasaka, and Takafumi Miyatake. Feature extraction of finger-vein patterns based on repeated line tracking and its application to personal identification. Machine Vision and Applications, 15(4):194–203, 2004.

[23] Annemarie Nadort. The hand vein pattern used as a biometric feature. Master Literature Thesis of Medical Natural Sciences at the Free University, Amsterdam, 2007.

[24] T.H.L.I.P.O. Nakazawa and K.H.L.I.P.O. Yazumi. Biometric information registration apparatus and method, December 18 2013. EP Patent 1,744,264.

[25] Madhuri M Pal and RW Jasutkar. Implementation of hand vein structure authentication based system. In Communication Systems and Network Technologies (CSNT), 2012 International Conference on, pages 114–118. IEEE, 2012.

[26] Chuchart Pintavirooj, Fernand S Cohen, and Woranut Lampa. Fingerprint alignment based on local feature combined with affine geometric invariant. 2010.

[27] R Deepak Prasanna, P Neelamegam, S Sriram, and Nagarajan Raju. Enhancement of vein patterns in hand image for biometric and biomedical application using various image enhancement techniques. Procedia Engineering, 38:1174–1185, 2012.

[28] M Rajalakshmi, R Rengaraj, and PG Student. Biometric authentication using near infrared images of palm dorsal vein patterns. International Journal of Advanced Engineering Technology, IJAET, 2, 2011.

[29] Debyo Saptono. Conception d'un outil de prototypage rapide sur le FPGA pour des applications de traitement d'images. PhD thesis, Dijon, 2011.

[30] R. Stewart. Security enhanced blasting apparatus with biometric analyzer and method of blasting, August 17 2011. EP Patent App. EP20,110,165,738.

[31] Chuan Chin Teo and Hong Tat Ewe. An efficient one-dimensional fractal analysis for iris recognition. 2005.

[32] Lingyu Wang and Graham Leedham. A thermal hand vein pattern verification system. In Pattern Recognition and Image Analysis, pages 58–65. Springer, 2005.

[33] Kejun Wang, Yan Zhang, Zhi Yuan, and Dayan Zhuang. Hand vein recognition based on multi supplemental features of multi-classifier fusion decision. In Mechatronics and Automation, Proceedings of the 2006 IEEE International Conference on, pages 1790–1795. IEEE, 2006.

[34] Lingyu Wang and Graham Leedham. Near-and far-infrared imaging for vein pattern biometrics. In Video and Signal Based Surveillance, 2006. AVSS'06. IEEE International Conference on, pages 52–52. IEEE, 2006.

[35] Leedham Wang, G Leedham, and S-Y Cho. Infrared imaging of hand vein patterns for biometric purposes. IET computer vision, 1(3):113–122, 2007.

[36] Lingyu Wang, Graham Leedham, and David Siu-Yeung Cho. Minutiae feature analysis for infrared hand vein pattern biometrics. Pattern recognition, 41(3):920–929, 2008.

[37] Jian-Gang Wang, Wei-Yun Yau, Andy Suwandy, and Eric Sung. Person recognition by fusing palmprint and palm vein images based on â€œlaplacianpalmâ€ representation. Pattern Recognition, 41(5):1514–1527, 2008.

[38] Meng-Hui Wang and Yu-Kuo Chung. Applications of thermal image and extension theory to biometric personal

recognition. Expert Systems with Applications, 39(8):7132–7137, 2012.

[39] Li Xueyan and Guo Shuxu. The Fourth Biometric-Vein Recognition. INTECH Open Access Publisher, 2008.

[40] Aycan Yuksel, Lale Akarun, and Bulent Sankur. Hand vein biometry based on geometry and appearance methods. IET computer vision, 5(6):398–406, 2011.

[41] Monwar, M., Rezaei, S., & Prkachin, K. Eigenimage based pain expression recognition. IAENG International Journal of Applied Mathematics, 36(2), 1-6. (2007).

[42] Jagadeesan, A., Thillaikkarasi, T., & Duraiswamy, K. Cryptographic key generation from multiple biometric modalities: Fusing minutiae with iris feature. International Journal of Computer Applications, 2(6), 16-26. (2010).

[43] Tanaka, Toshiyuki, and Naohiko Kubo. "Biometric authentication by hand vein patterns." SICE 2004 Annual Conference. Vol. 1. IEEE, 2004.

[44] Mostayed, Ahmed, et al. "Biometric authentication from low resolution hand images using radon transform." Computers and Information Technology, 2009. ICCIT'09. 12th International Conference on. IEEE, 2009.

[45] Lin, Lan, Yang Cong, and Yandong Tang. "Hand gesture recognition using RGB-D cues." Information and Automation (ICIA), 2012 International Conference on. IEEE, 2012.

[46] Sasikala, V. Bee Swarm based Feature Selection for Fake and Real Fingerprint Classification using Neural Network Classifiers. IAENG International Journal of Computer Science 42.4 (2015).

[47] Medjahed, Seyyid Ahmed, et al., Binary Cuckoo Search Algorithm for Band Selection in Hyperspectral Image Classification, IAENG International Journal of Computer Science 42.3 (2015).

Biography

AUTHORS PROFILE

Sarâh BENZIANE is assistant professor in computer science; she obtained her magister electronics about mobile robotics. She holds a basic degree from computer science engineering. Now, she’s working with biometrics system’s processing in SMPA laboratory, at the university of Science and Technology of Oran Mohamed Boudiaf (Algeria). She teaches at the University of Oran at the Maintenance and Industrial Safety Institute. Her current research interests are in the area of artificial intelligence and image processing, mobile robotics, neural networks, Biometrics, neuro-computing, GIS and system engineering.

Abdelkader Benyettou received the engineering degree in 1982 from the Institute of Telecommunications of Oran and the MSc degree in 1986 from the University of Sciences and Technology of Oran-USTO, Algeria. In 1987, he joined the Computer Sciences Research Center of Nancy, France, where he worked until 1991 on Arabic speech recognition by expert systems (ARABEX) and received the PhD in electrical engineering in 1993 from the USTOran University. From 1988 to 1990, he has been an assistant Professor in the department of Computer Sciences, Metz University, and Nancy-I University. He is actually professor at the USTOran University since 2003. He is currently a researcher director of the Signal-Speech-Image– SIMPA Laboratory, department of Computer Sciences, Faculty of sciences, USTOran, since 2002. His current research interests are in the area of speech and image processing, automatic speech recognition, neural networks, artificial immune systems, genetic algorithms, neuro-computing, machine learning, neuro-fuzzy logic, handwriting recognition, electronic/electrical engineering, signal and system engineering.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Blind image separation based on exponentiated transmuted Weibull distribution

A. M. Adam
Department of Mathematics, Faculty of Science,
Zagazig University, P.O. Box
Zagazig, Egypt.

R. M. Farouk
Department of Mathematics, Faculty of Science,
Zagazig University, P.O. Box
Zagazig, Egypt.

M. E. Abd El-aziz
Department of Mathematics, Faculty of Science,
Zagazig University, P.O. Box
Zagazig, Egypt

***Abstract*-In recent years the processing of blind image separation has been investigated. As a result, a number of feature extraction algorithms for direct application of such image structures have been developed. For example, separation of mixed fingerprints found in any crime scene, in which a mixture of two or more fingerprints may be obtained, for identification, we have to separate them. In this paper, we have proposed a new technique for separating a multiple mixed images based on exponentiated transmuted Weibull distribution. To adaptively estimate the parameters of such score functions, an efficient method based on maximum likelihood and genetic algorithm will be used. We also calculate the accuracy of this proposed distribution and compare the algorithmic performance using the efficient approach with other previous generalized distributions. We find from the numerical results that the proposed distribution has flexibility and an efficient result.**

Keywords- Blind image separation, Exponentiated transmuted Weibull distribution, Maximum likelihood, Genetic algorithm, Source separation, FastICA.

I. INTRODUCTION

Recently the blind source separation (BSS) has more attention because it can be considered as an advanced image/signal processing technique and has many applications such as: speech sound, image, communication, and biomedicine [1–4]. BSS aims to recover source (images/signals) from a mixture with little known information. There are many BSS algorithms that have been discussed from various viewpoints, including principle component analysis (PCA) [9], maximum likelihood [7], mutual information minimization [6], tensors [8], non-Gaussianity [5], and neural networks [10-12]. Regarding to BSS, the separation and optimization methods play the most important roles. Separation step is used as the measurement of separability and optimization step is used to get the optimum solution for the objective function which we get from separation mechanism. Using generalized distributions usually gives good results of blind separation due to the variant properties of its sub-models. In the independent component analysis (ICA) framework, accurately estimates the statistical model of the sources is still an open and challenging problem [2]. Practical BSS scenarios employ difficult source distributions and even situations where many sources with variant probability density functions (pdf) mixed together. Towards this direction, many parametric density models have been made available in recent literature. For examples of such models, the generalized Gaussian density

(GGD) [13], the generalized gamma density (GGD) [14], and even combinations and generalizations such as super and generalized Gaussian mixture model (GMM) [15], the Pearson family of distributions [16], the generalized alfa-beta distribution (AB-divergences) [17] and even the so-called extended generalized lambda distribution (EGLD) [18] which is an extended parameterizations of the aforementioned generalized lambda distribution (GLD) and generalized beta distribution (GBD) models [19]. In this paper, we have presented the exponentiated transmuted Weibull distribution (ETWD) which is a generalization of the Weibull distribution. We have evaluated the accuracy of our proposed ETWD and compare the algorithmic performance using many different previous distributions. The numerical results, shows that the ETWD give a good results comparing with many different cases. The rest of this paper is organized as follows: In section 2, we present the BSS model. In section 3, we will discuss the ETWD. In section 4, we will use maximum likelihood to estimate the parameters of ETWD based on genetic algorithm. Finally, we will present the computational efficient performance of our proposed technique.

II. BLIND SOURCE SEPARATION (BSS) MODEL

Let $S(t) = [s_1(t), s_2(t), \dots, s_N(t)]^T (t = 1, 2, \dots, l)$ denote independent source image vector that comes from N image sources. We can get observed mixtures

$X(t) = [x_1(t), x_2(t), \dots, x_K(t)]^T (N = K)$ under the circumstances of instantaneous linear mixture.

$$X(t) = AS(t), \tag{1}$$

where $\mathbf{A}$ is a $N \times N$ mixing matrix. The task of the BSS algorithm is to recover the sources from mixtures $x(t)$ by using

$$U(t) = WX(t), \tag{2}$$

where $\mathbf{W}$ is a $N \times N$ separation matrix and $U(t) = [u_1(t), u_2(t), \dots, u_N(t)]^T$ is the estimate of N sources.

Often sources are assumed to be zero-mean and unit-variance signals with at most one having a Gaussian distribution. To solve the problem of source estimation the un-mixing matrix W must be determined. In general, the majority of BSS approaches perform ICA, by essentially optimizing the negative log-likelihood (objective) function with respect to the un-mixing matrix W such that

$$L(u, W) = \sum_{l=1}^{N} E[\log p_{ul}(u_l)] - \log|\det(W)|, \tag{3}$$

where $E[.]$ represents the expectation operator and $p_{u1}(u_1)$ is the model for the marginal pdf of u_l, for all $l = 1, 2, \dots, N$. In effect, when correctly hypothesizing upon the distribution of the sources, the maximum likelihood (ML) principle leads to estimating functions, which in fact are the score functions of the sources

$$\varphi_l(u_l) = -\frac{d}{du_l}\log p_{ul}(u_l) \tag{4}$$

In principle, the separation criterion in (3) can be optimized by any suitable ICA algorithm where contrasts are utilized (see; e.g., [2]). The FastICA [3], based on

$$W_{k+1} = W_k + D(E[\varphi(u)u^T] - \mathrm{diag}(E[\varphi_l(u_l)u_l]))W_k\,, \tag{5}$$

where, as defined in [4],

$$D = \mathrm{diag}\left(\frac{1}{E[\varphi_l(u_l)u_l] - E[\varphi_l'(u_l)]}\right), \tag{6}$$

where $\varphi(t) = [\varphi_1(u_1), \varphi_2(u_2), \dots, \varphi_n(u_n)]^T$, valid for all $l = 1, 2, \dots, n$.

In the following section, we propose ETWD for image modeling.

III. Exponentiated transmuted Weibull distribution (ETWD)

Following [20] ETWD is a new generalization of the two parameters Weibull distribution. The pdf of ETWD is defined as:

$$f(x) = \frac{\nu\beta}{\alpha}\left(\frac{x_i}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x_i}{\alpha}\right)^{\beta}}\left[1-\lambda+2\lambda e^{-\left(\frac{x_i}{\alpha}\right)^{\beta}}\right]\times\left[1+(\lambda-1)e^{-\left(\frac{x_i}{\alpha}\right)^{\beta}} - \lambda e^{-2\left(\frac{x_i}{\alpha}\right)^{\beta}}\right]^{\nu-1} \tag{7}$$

cumulative distribution function of ETWD is given by:

$$F(x) = \left\{1+(\lambda-1)e^{-(\frac{x}{\alpha})^{\beta}} - \lambda e^{-2(\frac{x}{\alpha})^{\beta}}\right\}^{\nu} x \geq 0\,, \tag{8}$$

where $\alpha, \beta > 0$, and $|\lambda| \leq 1$ are the scale, shape and transmuted parameters, respectively. It is clear that the ETWD is very flexible. This is so since there are many several other distributions that can be considered as special cases of ETW, by selecting the appropriate values of the parameters. These special cases include eleven distributions as shown in Table (I). In Figure (1-4) there are several distributions generated from ETWD by changing the parameters.

IV. Estimation of the parameters

To estimate the parameters of ETWD, the maximum likelihood is used.

Let $X_1, X_2 \dots, X_n$ be a sample of size N from an ETWD.

Then the log-likelihood function ($\mathcal{L}$) is given by:

$$\mathcal{L} = \log \ell = \log\left(\prod_{i=1}^{n}\left[\frac{\nu\beta}{\alpha}\left(\frac{x_i}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x_i}{\alpha}\right)^{\beta}}\left[1-\lambda+2\lambda e^{-\left(\frac{x_i}{\alpha}\right)^{\beta}}\right]\times\left[1+(\lambda-1)e^{-\left(\frac{x_i}{\alpha}\right)^{\beta}} - \lambda e^{-2\left(\frac{x_i}{\alpha}\right)^{\beta}}\right]^{\nu-1}\right]\right) \tag{9}$$

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Therefore, maximum likelihood estimation of α, β, λ and ν are derived from the derivatives of $\mathcal{L}$. They should satisfy the following equations: $\frac{\partial\mathcal{L}}{\partial\alpha}=0,\ \frac{\partial\mathcal{L}}{\partial\lambda}=0\ ,\frac{\partial\mathcal{L}}{\partial\beta}=0\ ,\frac{\partial\mathcal{L}}{\partial\nu}=0$

$$\frac{\partial\mathcal{L}}{\partial\alpha}=-\frac{n\beta}{\alpha}+\beta\sum_{i=1}^{n}\left(\frac{x_i}{\alpha}\right)^{\beta-1}+\sum_{i=1}^{n}\frac{2\lambda e^{-\left(\frac{x_i}{\alpha}\right)^{\beta}}\beta\left(\frac{x_i}{\alpha}\right)^{\beta-1}\left(\frac{x_i}{\alpha^2}\right)}{2\lambda e^{-\left(\frac{x_i}{\alpha}\right)^{\beta}}-\lambda+1}+(\nu-1)$$
$$\times\sum_{i=1}^{n}\frac{(\lambda-1)e^{-\left(\frac{x_i}{\alpha}\right)^{\beta}}\beta\left(\frac{x_i}{\alpha}\right)^{\beta-1}\left(\frac{x_i}{\alpha^2}\right)-2\lambda e^{-2\left(\frac{x_i}{\alpha}\right)^{\beta}}\beta\left(\frac{x_i}{\alpha}\right)^{\beta-1}\left(\frac{x_i}{\alpha^2}\right)}{(\lambda-1)e^{-\left(\frac{x_i}{\alpha}\right)^{\beta}}-\lambda e^{-2\left(\frac{x_i}{\alpha}\right)^{\beta}}+1} \tag{10}$$

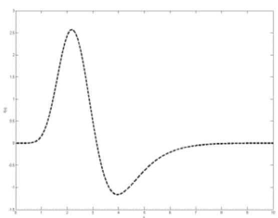
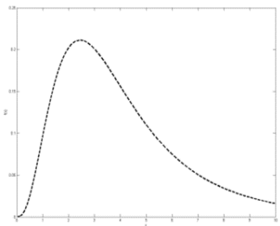
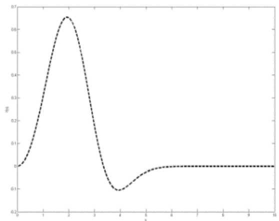
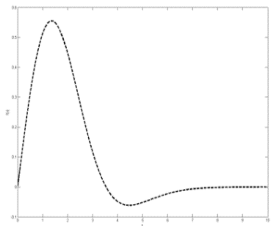
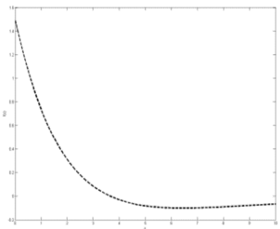
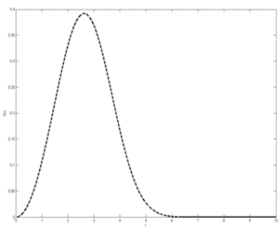
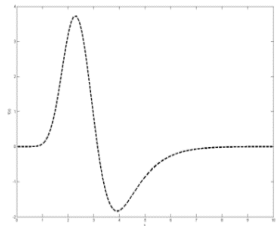
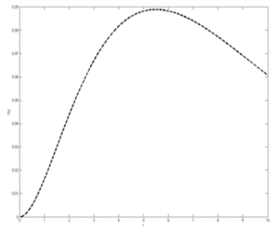
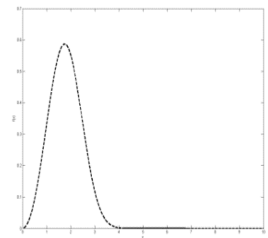
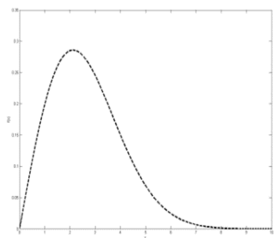
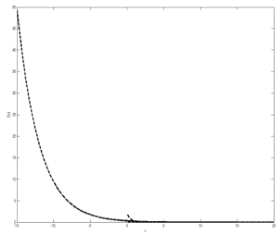
$$\frac{\partial\mathcal{L}}{\partial\beta}=\frac{n}{\beta}+\sum_{i=1}^{n}\log(x_i)-n\log\alpha-\sum_{i=1}^{n}\left(\frac{x_i}{\alpha}\right)^{\beta}\log\left(\frac{x_i}{\alpha}\right)+\sum_{i=1}^{n}\frac{-2\lambda e^{-\left(\frac{x}{\alpha}\right)^{\beta}}\left(\frac{x}{\alpha}\right)^{\beta}\log\left(\frac{x_i}{\alpha}\right)}{2\lambda e^{-\left(\frac{x_i}{\alpha}\right)^{\beta}}-\lambda+1}+(\nu-1)$$
$$\times\sum_{i=1}^{n}\frac{-(\lambda-1)e^{-\left(\frac{x_i}{\alpha}\right)^{\beta}}\left(\frac{x_i}{\alpha}\right)^{\beta}\log\left(\frac{x_i}{\alpha}\right)+2\lambda e^{-2\left(\frac{x_i}{\alpha}\right)^{\beta}}\left(\frac{x_i}{\alpha}\right)^{\beta}\log\left(\frac{x_i}{\alpha}\right)}{(\lambda-1)e^{-\left(\frac{x_i}{\alpha}\right)^{\beta}}-\lambda e^{-2\left(\frac{x_i}{\alpha}\right)^{\beta}}+1} \tag{11}$$

$$\frac{\partial\mathcal{L}}{\partial\lambda}=\sum_{i=1}^{n}\frac{2e^{-\left(\frac{x_i}{\alpha}\right)^{\beta}}-1}{2\lambda e^{-\left(\frac{x_i}{\alpha}\right)^{\beta}}-\lambda+1}+(\nu-1)\times\sum_{i=1}^{n}\frac{e^{-\left(\frac{x_i}{\alpha}\right)^{\beta}}-e^{-2\left(\frac{x_i}{\alpha}\right)^{\beta}}}{(\lambda-1)e^{-\left(\frac{x_i}{\alpha}\right)^{\beta}}-\lambda e^{-2\left(\frac{x_i}{\alpha}\right)^{\beta}}+1} \tag{12}$$

$$\frac{\partial\mathcal{L}}{\partial\nu}=\sum_{i=1}^{n}\log\left[(\lambda-1)e^{-\left(\frac{x_i}{\alpha}\right)^{\beta}}-\lambda e^{-2\left(\frac{x_i}{\alpha}\right)^{\beta}}+1\right]+\frac{n}{\nu} \tag{13}$$

To estimate the value of parameters, the system of equations (10-13) must be solved. However, it is difficult to solve this system so, the genetic algorithm (GA) [21-22] will be used as an alternative numerical method to estimate the parameters. The appeal of the GA optimization technique lies in the fact that it can minimize the negative of the log-likelihood objective function in (3), essentially without depending on any derivative information.

Table I
The ETWD sub-models, shows the specific values of the parameters used to generate the above mentioned eleven special cases, Where $\alpha > 0$, $\beta > 0$, $\nu > 0$, $|\lambda| \leq 1$

 $\beta = 2$ Exponentiated transmuted Rayleigh (ETR)	 $\beta = 1$ Exponentiated transmuted exponential (ETE)	 $\nu = 1$ Transmuted Weibull (TW)	 $\beta = 2, \nu = 1$ Transmuted Rayleigh (TR)
 $\beta = 1, \nu = 1$ Transmuted exponential (TE)	 $\lambda = 0$ Exponentiated Weibull (EW)	 $\beta = 2, \lambda = 0$ Exponentiated Rayleigh (ER)	 $\beta = 1, \lambda = 0$ Exponentiated exponential (EE)
 $\lambda = 0, \nu = 1$ Weibull (W)	 $\beta = 2, \lambda = 0, \nu = 1$ Rayleigh (R)	 $\beta = 1, \lambda = 0, \nu = 1$ Exponential (E)	

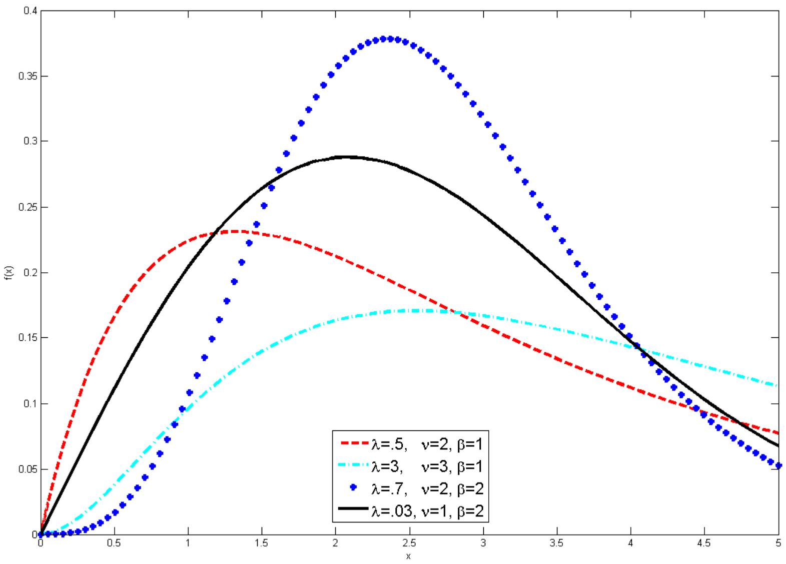


Figure 1. The ETWD with fixed α=3.

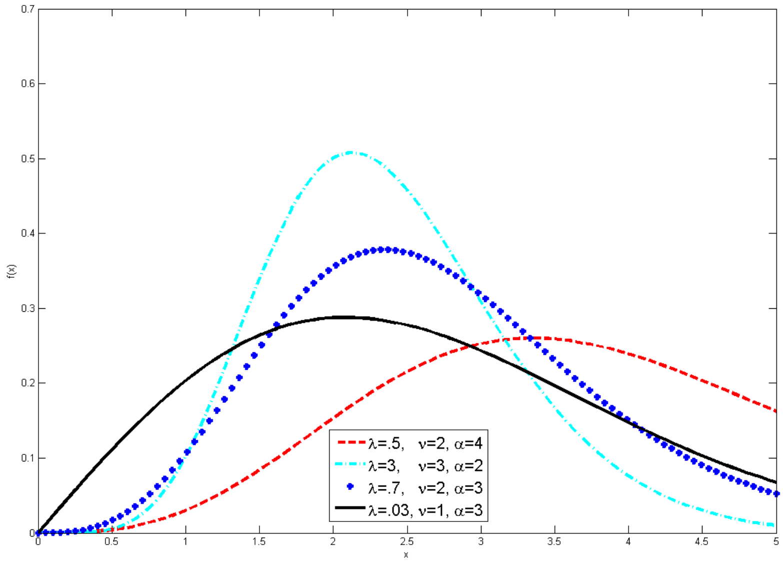


Figure 2. The ETWD with fixed **β**=2.

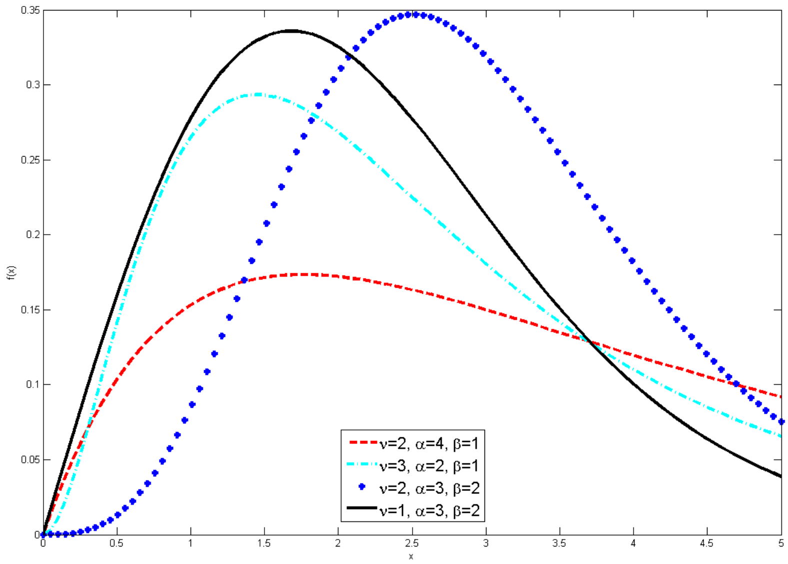


Figure 3. The ETWD with fixed λ=0.5.

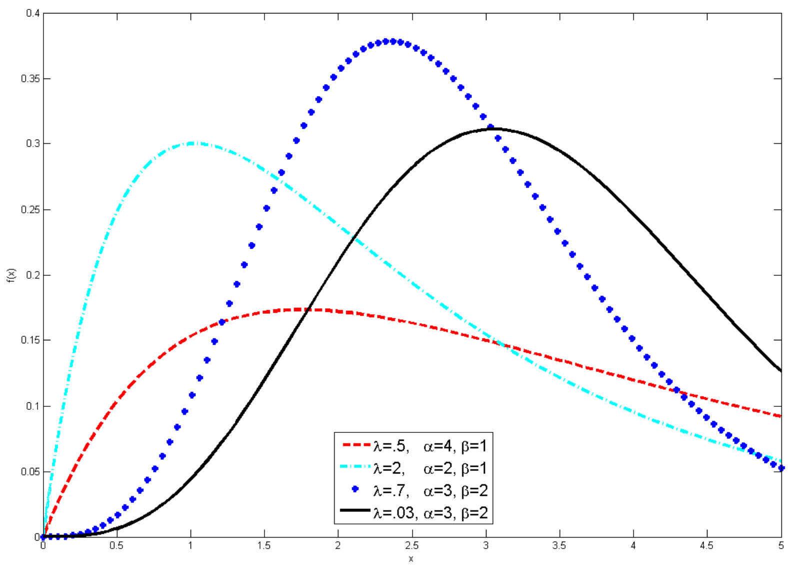


Figure 4. The ETWD with fixed ν=2.

V. NUMERICAL RESULTS

Numerical experiments have shown that the GA method can converge to an acceptably accurate solution with substantially fewer function evaluations. We have generated random number from ETWD with parameters α, β, ν and λ. By performing GA, we obtain best estimation of parameters as in table (II).

Applications of ETWD for BSS

We resolve to FastICA algorithm for blind signal separation (BSS). This algorithm depends on the estimated parameters and an un-mixing matrix W which estimated by FastICA algorithm. By substituting (7) into (4) for the source estimates u_l, $l = 1, 2, \dots, n$, it quickly becomes clear that the proposed score function inherits a generalized parametric structure, which can be attributed to the highly flexible ETWD parent model. So, a simple calculus yields the flexible BSS score function

$$\varphi_l(u_l) = -\frac{d}{du_l}\log\frac{\nu\beta}{\alpha}\left(\frac{x_i}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x_i}{\alpha}\right)^{\beta}}\left[1-\lambda+2\lambda e^{-\left(\frac{x_i}{\alpha}\right)^{\beta}}\right]\times\left[1+(\lambda-1)e^{-\left(\frac{x_i}{\alpha}\right)^{\beta}}-\lambda e^{-2\left(\frac{x_i}{\alpha}\right)^{\beta}}\right]^{\nu-1} \quad (14)$$

In principle $\varphi_l(u_l|\theta)$ is capable of modeling a large number of signals as well as various other types of challenging heavy- and light-tailed distributions. Experiments were done to investigate the performance of our method through three applications (two in source separation and one in image denoising) when impulsive noise is presented. In all experiments, the performance of our method is compared with generalized gamma [14], tanh, skew, pow3 [23], and Gauss [15]. Our performance is measured by the peak-signal-to- noise ratio (PSNR), defined as:

$$PSNR = 20\, log_{10}\left(\frac{255}{MSE}\right) \quad (15)$$

Table II
Parameter estimation by using GA

	λ	ν	α	β	$\hat{\lambda}$	$\hat{\nu}$	$\hat{\alpha}$	$\hat{\beta}$	Err
X1	0.5	2	3	4	0.59	1.86	2.97	4.11	0.02
X2	1	2.5	5.2	6.8	1.16	2.42	5.27	6.80	0.06
X3	3	5.7	1.9	8.2	2.98	5.63	1.98	8.12	0.006

Example **1**

We have run the algorithm using natural images taken from [24]. We selected 4 noise-free natural images with 512×512 pixels. Further, to reduce the dimension of input image data, the data set X is centered and whitened by principal component analysis (PCA) method. Then, using the updating rules of W defined in (5), the objective function given in (14) is minimized. Where Figure (5-6) show the original, mixed and separated images by Gauss, pow3, skew, tanh, generalized gamma, and ETWD algorithms. Also, Table (III) illustrates the performance of these algorithms. From this table and Figure (5-6), the ETWD is higher performance than other algorithms.

Table III
Image separation PSNR

Distribution / PSNR	First Image		Second Image		Third Image		Forth Image		Elapsed time (in seconds)
	MSE	PSNR	MSE	PSNR	MSE	PSNR	MSE	PSNR	
Gauss	0.1176	57.4255	0.2972	53.4009	0.1773	55.6426	0.1314	56.9444	8.757703
Pow3	0.1375	56.7477	0.2130	54.8477	0.1736	55.7363	0.1259	57.1320	24.921161
Skew	0.0044	71.7366	0.0177	65.6481	0.2340	54.4378	0.2193	54.7209	5.788523
Tanh	0.1179	57.4172	0.1647	55.9628	0.1810	55.5538	0.0741	59.4309	6.852007
Generalized Gamma	0.1341	56.8571	0.2659	53.8840	0.1865	55.4237	0.1305	56.9746	4.333974
ETWD	0.0011	77.6298	0.0159	66.1132	0.0026	73.9429	0.0015	76.2714	4.285013

Example **2**

In this example, we illustrate the performance of our algorithm to denoise medical images taken from [25]. Where Figure (7-12) show the original images, noised images, and denoised images by different algorithms. After applying algorithms of Gauss, pow3, skew, tanh, generalized gamma and, our algorithm ETWD, the results are illustrated in Figure (7- 12), also Table (IV) illustrates the performance of these algorithms. From table (IV) and Figure (7-12), the ETWD is higher performance than other algorithms.

Table IV
Denoising PSNR

Distribution / PSNR	First Image (Medical)		Second Image (Medical)		Elapsed time (in seconds)
	MSE	PSNR	MSE	PSNR	
Gauss	0.0092	68.4753	0.0077	69.2751	1.724821
Pow3	0.0077	69.2780	0.0093	68.4489	1.646659
Skew	0.0077	69.2797	0.0093	68.4383	1.611382
Tanh	0.0076	69.2967	0.0093	68.4483	1.729392
Generalized gamma	0.0058	70.5134	0.0061	70.2859	1.578206
ETWD	0.0050	71.1162	0.0039	72.1719	1.646362

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

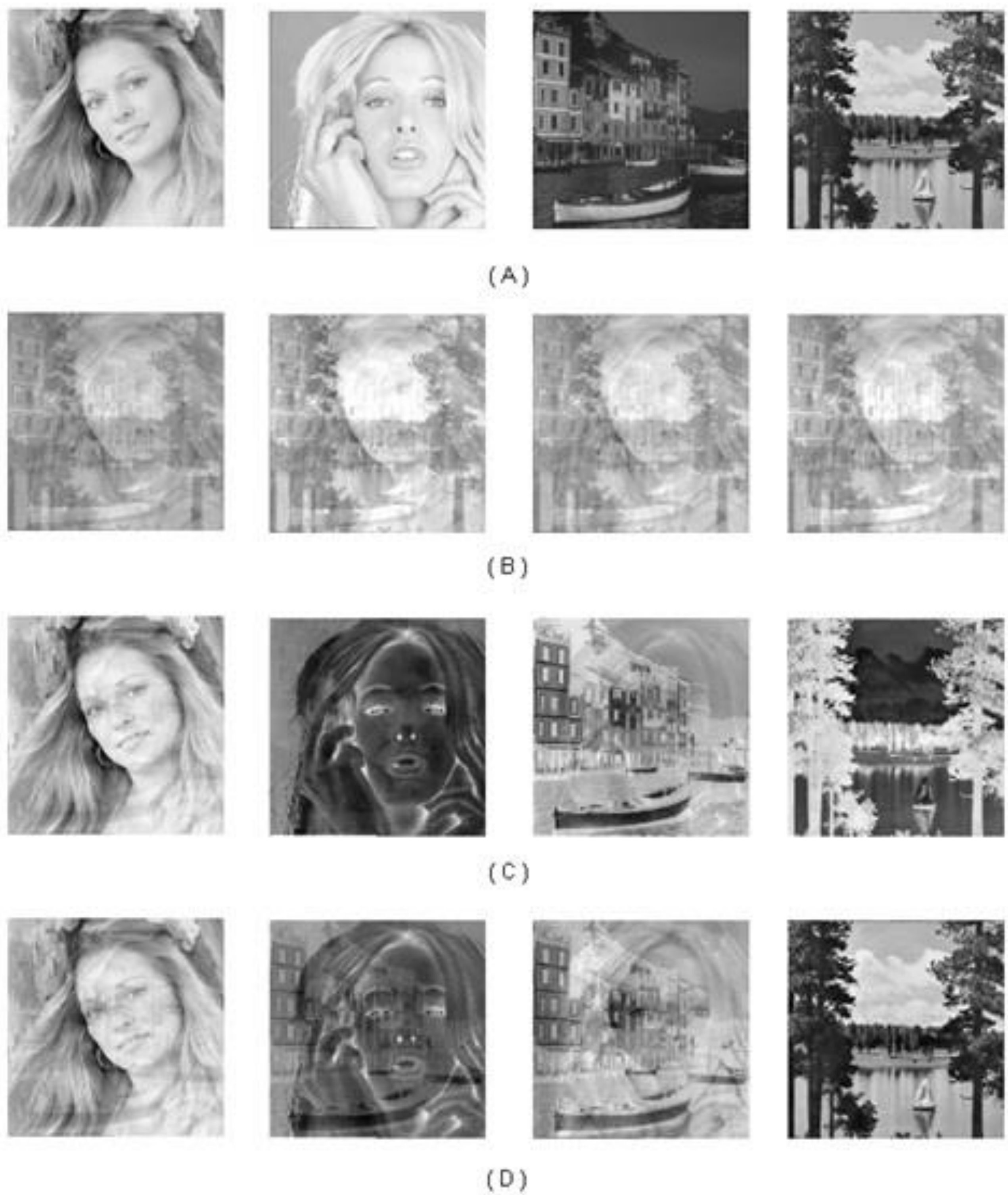


Figure 5. A original images, B mixed images, C Gauss separated images, and D pow3 separated images.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

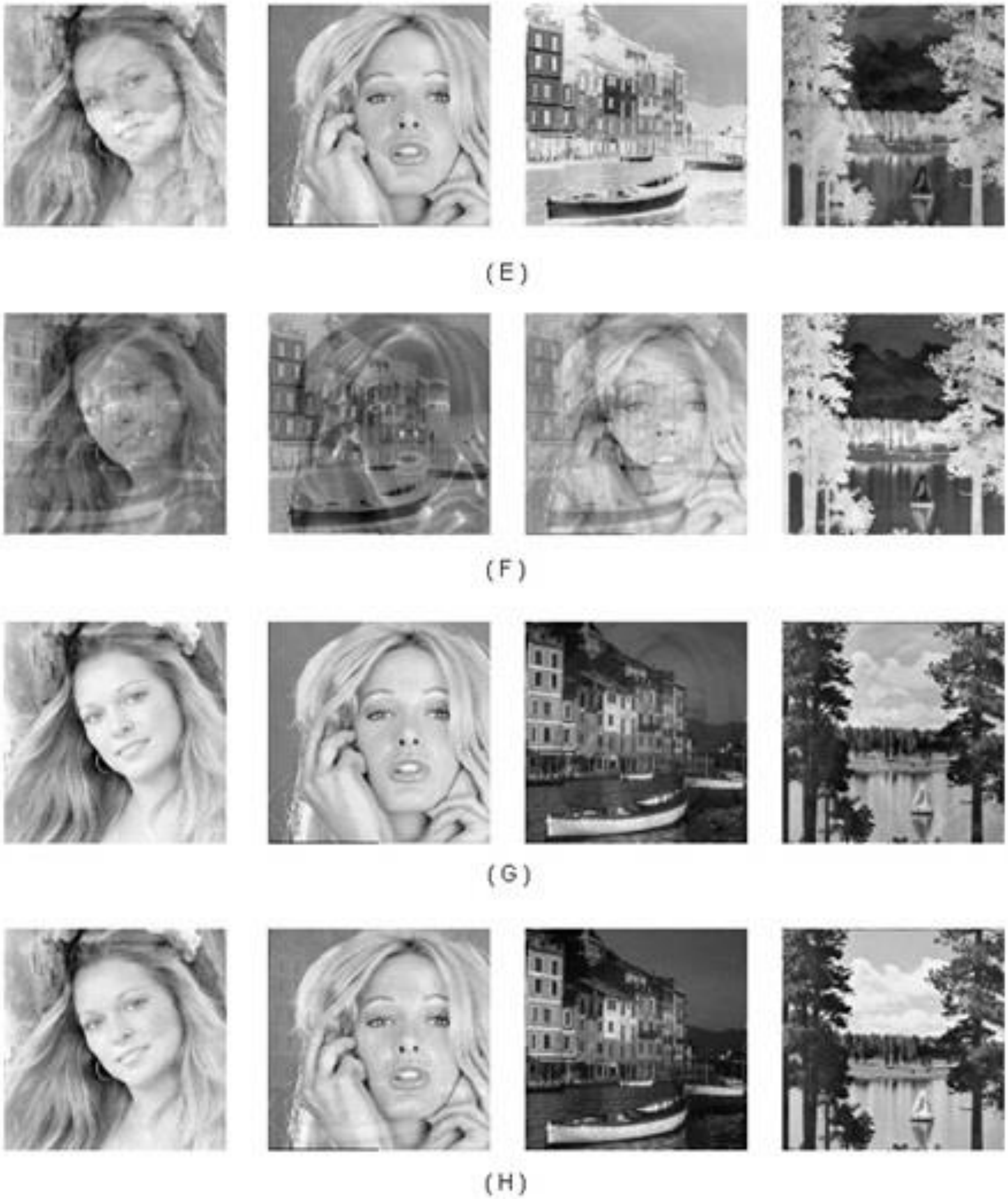


Figure 6. E skew separated images, F tanh separated images, G generalized gamma separated images, and H ETWD separated images.

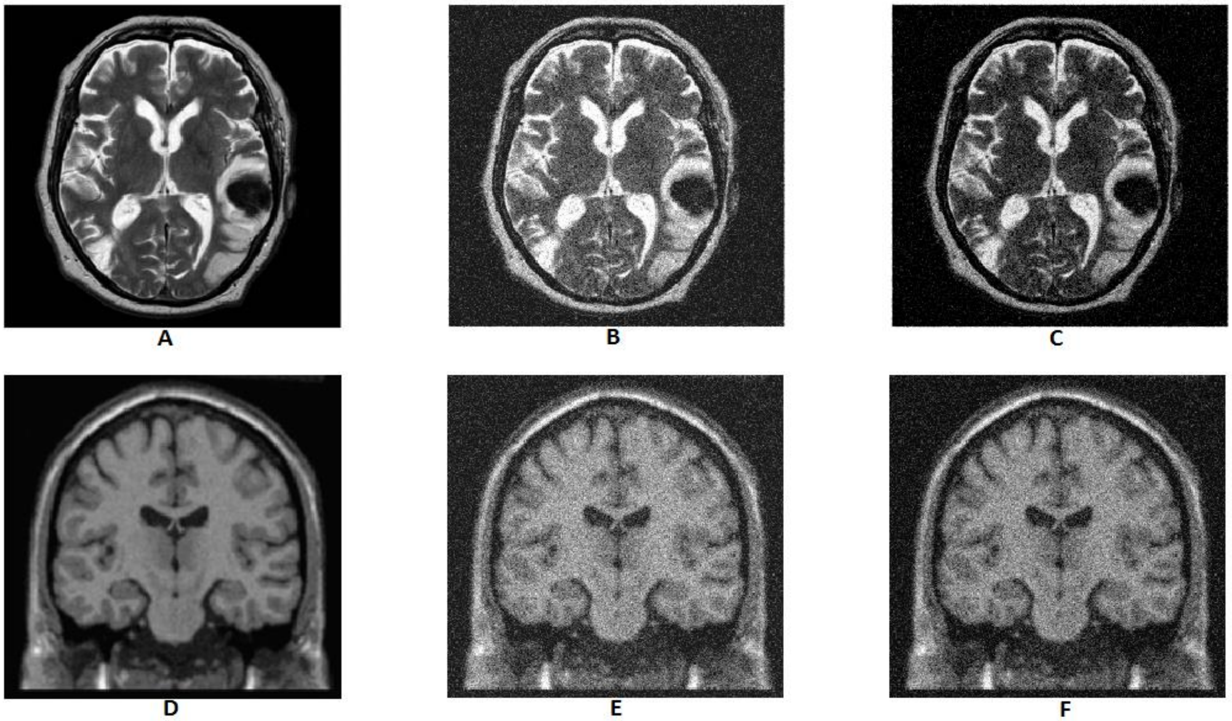


Figure 7. Medical image denoising using Gauss filter: A, D are the source images, B, E are the noised images, C, F are the denoised images.

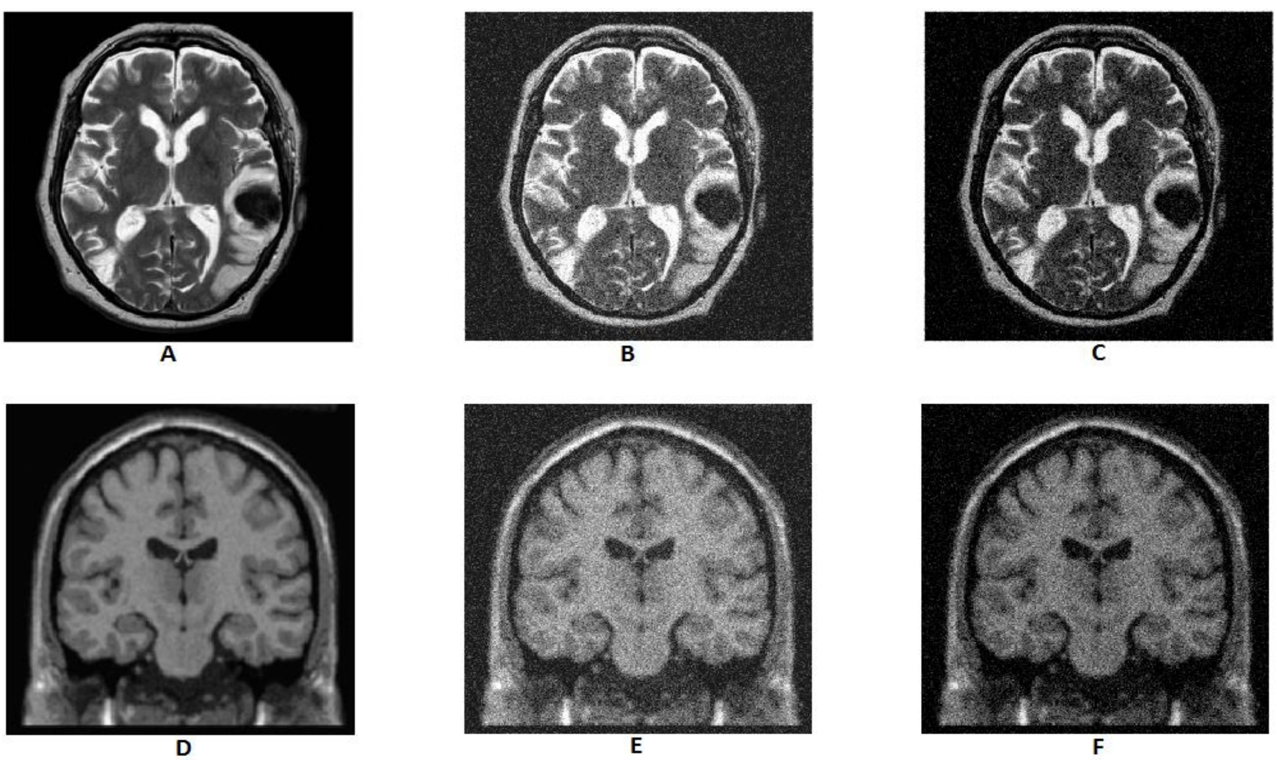


Figure 8. Medical image denoising using pow3 filter: A, D are the source images, B, E are the noised images, C, F are the denoised images.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

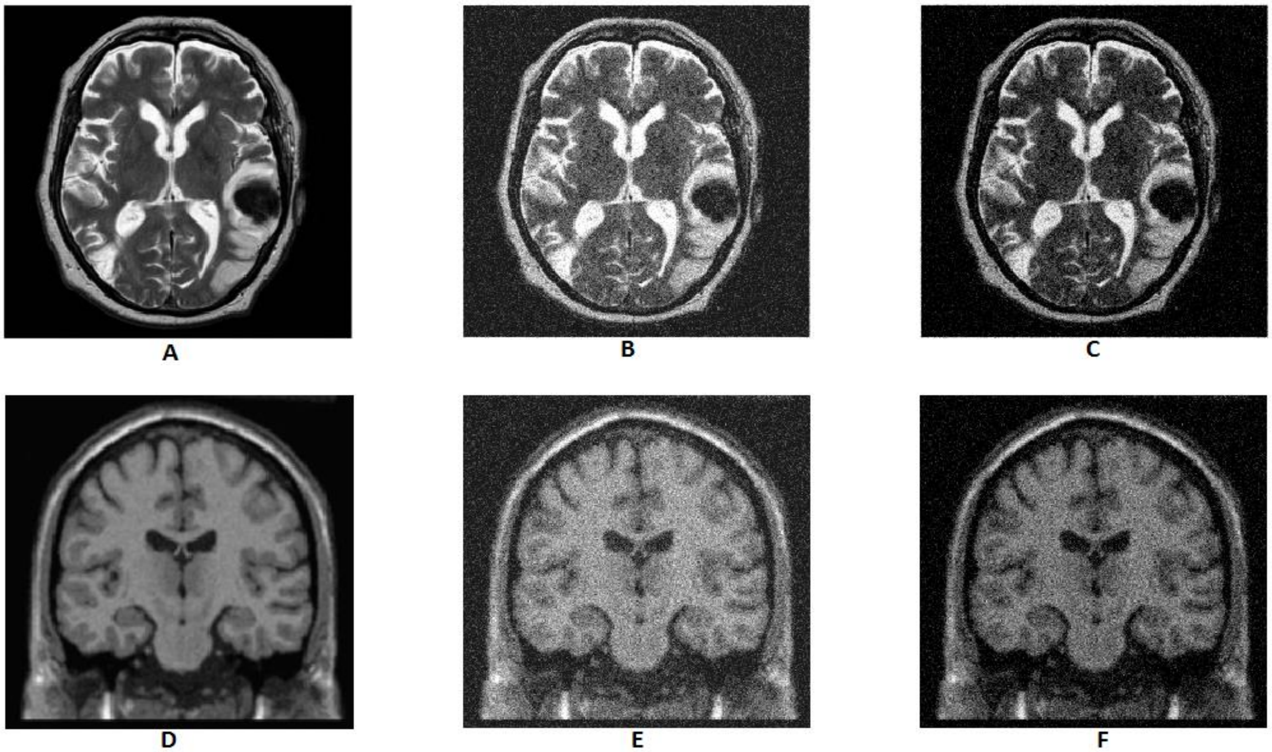


Figure 9. Medical image denoising using Skew filter: A, D are the source images, B, E are the noised images, C, F are the denoised images.

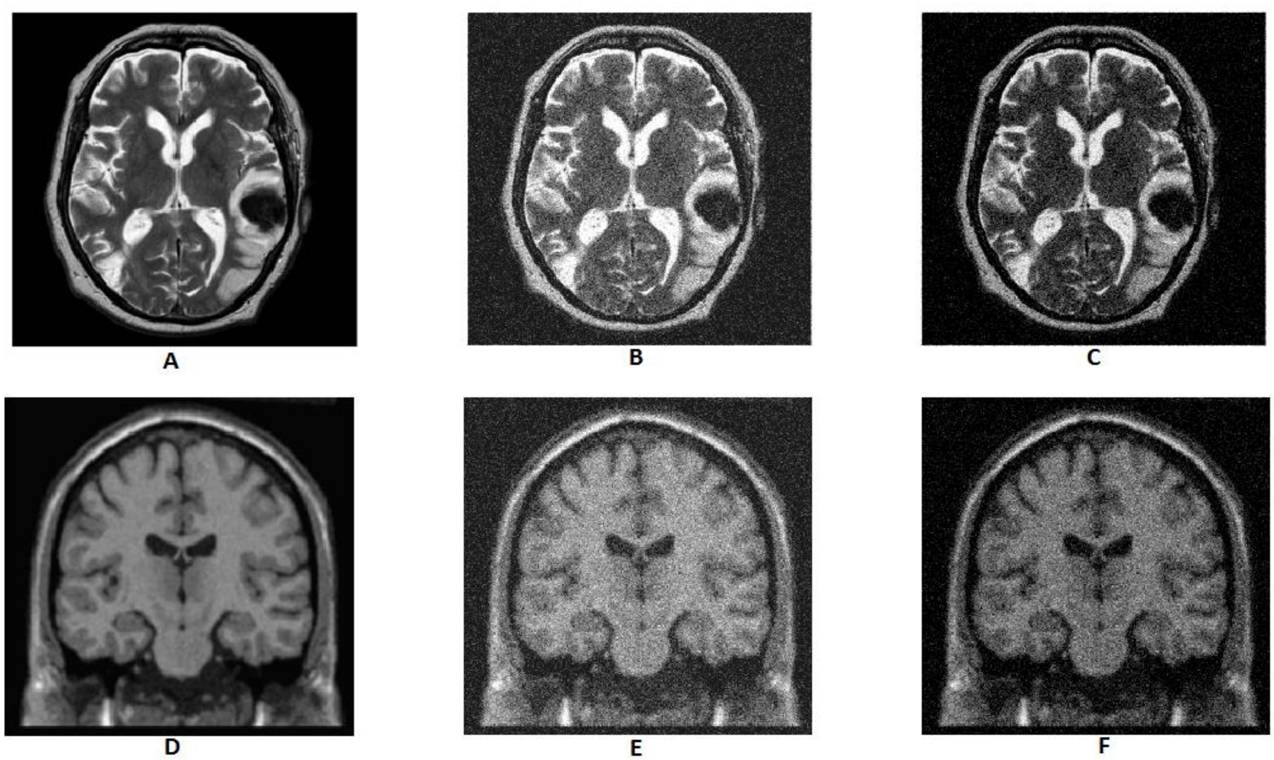


Figure 10. Medical image denoising using tanh filter: A, D are the source images, B, E are the noised images, C, F are the denoised images.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

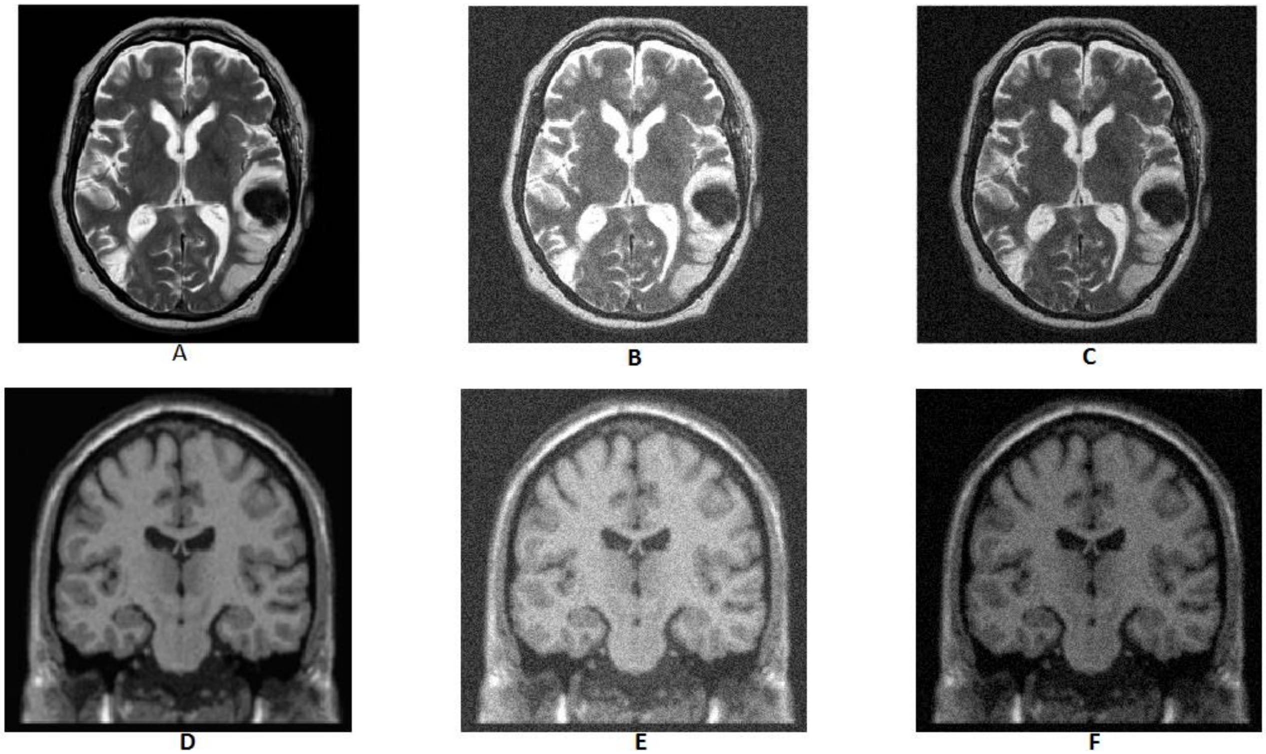


Figure 11. Medical image denoising using generalized gamma filter: A, D are the source images, B, E are the noised images, C, F are the denoised images.

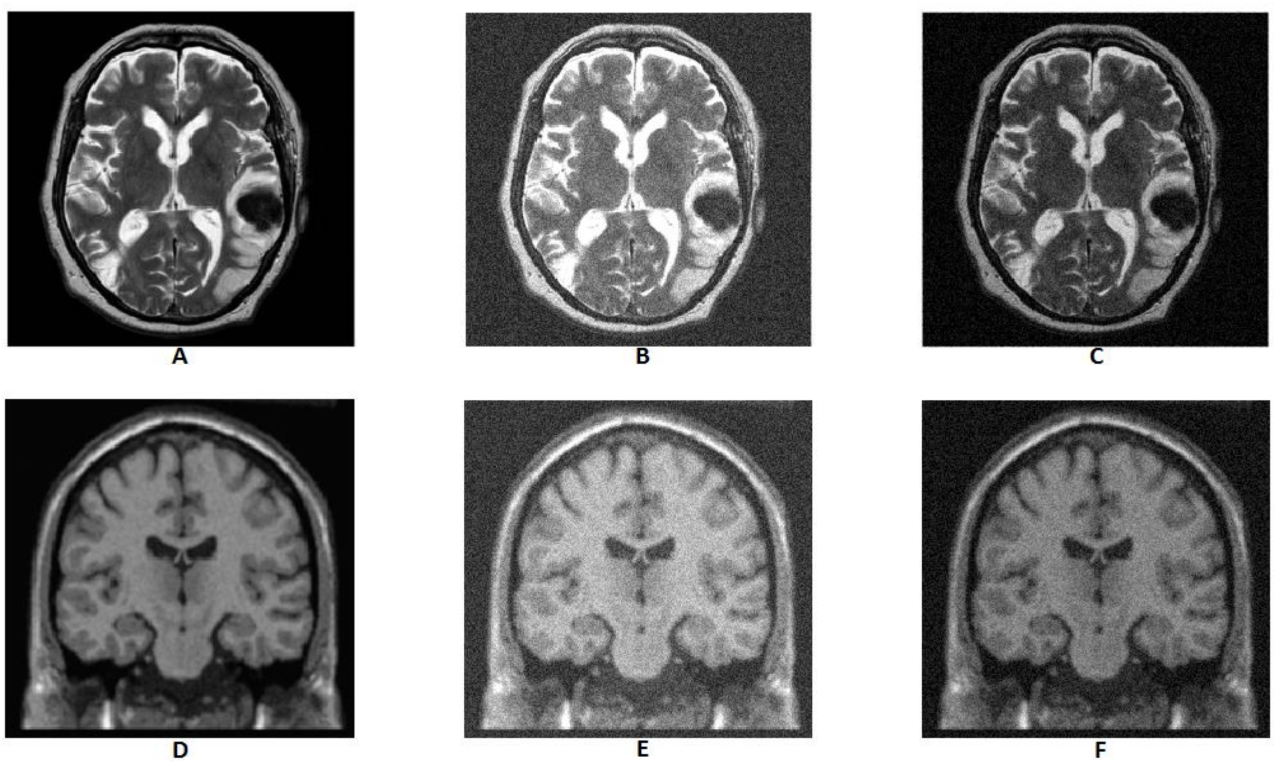


Figure 12. Medical image denoising using ETWD filter: A, D are the source images, B, E are the noised images, C, F are the denoised images.

VI. CONCLUSION

In this paper, we introduced a new technique for blind image separation and image denoise based on exponentiated transmuted Weibull distribution. Our proposed technique outperforms existing solutions in terms of separation quality and computational cost. When the GA is used to estimate the parameters of ETWD and it gives small error. Also the results of ETWD are better than other algorithms.

REFERENCES

[1] Y. Zhang and Y. Zhao, 2013. Modulation domain blind speech separation in noisy environments, Speech Communication, vol. 55, no. 10, pp. 1081–1099.
[2] M. T. ¨ Ozgen, E. E. Kuruo˘glu, and D. Herranz, 2009. Astrophysical image separation by blind time-frequency source separation methods, Digital Signal Processing, vol. 19, no. 2, pp. 360–369.
[3] Ikhlef, K. Abed-Meraim, and D. Le Guennec, 2010. Blind signal separation and equalization with controlled delay for MIMO convolutive systems, Signal Processing, vol. 90, no. 9, pp. 2655– 2666.
[4] R. Romo V´azquez, H. V´elez-P´erez, R. Ranta, V. Louis Dorr, D. Maquin, and L. Maillard, Blind source separation, 2012. Wavelet denoising and discriminant analysis for EEG artifacts and noise cancelling, Biomedical Signal Processing and Control, vol. 7, no. 4, pp. 389–400.
[5] M. Kuraya, A. Uchida, S. Yoshimori, and K. Umeno, 2008. Blind source separation of chaotic laser signals by independent component analysis, Optics Express, vol. 16, no. 2, pp. 725–730.
[6] M. Babaie-Zadeh and C. Jutten, 2005. A general approach formutual information minimization and its application to blind source separation, Signal Processing, vol. 85, no. 5, pp. 975–995.
[7] K. Todros and J. Tabrikian, 2007. Blind separation of independent sources using Gaussian mixture model, IEEE Transactions on Signal Processing, vol. 55, no. 7, pp. 3645–3658.
[8] P. Comon, 2014. Tensors: a brief introduction, IEEE Signal Processing Magazine, vol. 31, no. 2, pp. 44–53.
[9] E. Oja and M. Plumbley, April 2003. Blind separation of positive sources using nonnegative PCA, in Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA '03), Nara, Japan, pp. 11–16.
[10] W. L. Woo and S. S. Dlay, 2005. Neural network approach to blind signal separation of mono-nonlinearly mixed sources, IEEE Transactions on Circuits and Systems I, vol. 52, no. 6, pp. 1236–1247.
[11] Cichocki and R. Unbehauen, 1996. Robust neural networks with on-line learning for blind identification and blind separation of sources, IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, vol. 43, no. 11, pp. 894–906.
[12] S.-I. Amari, T.-P. Chen, and A. Cichocki, 1997. Stability analysis of learning algorithms for blind source separation, Neural Networks, vol. 10, no. 8, pp. 1345–1351.
[13] K. Kokkinakis and A. K. Nandi, 2005. Exponent parameter estimation for generalized Gaussian probability density functions with application to speech modeling, Signal Processing, vol. 85, no. 9, pp. 1852–1858.
[14] E. W. Stacy, 1962. A generalization of the gamma distribution, Annals of Mathematical Statistics, vol. 33, no. 3, pp. 1187–1192.
[15] J. A. Palmer, K. Kreutz-Delgado, and S. Makeig, March 2006. Super-Gaussian mixture source model for ICA, in Proceedings of the International Conference on Independent Component Analysis and Blind Signal Separation, Charleston, SC, USA pp. 854–861.
[16] J. Eriksson, J. Karvanen, and V. Koivunen, 2002. Blind separation methods based on Pearson system and its extensions, Signal Processing, vol. 82, no. 4, pp. 663–673.
[17] Sarmiento, I. Durán-Díaz, A. Cichocki, and S. Cruces, 2015. A contrast based on generalized divergences for solving the permutation problem of convolved speech mixtures, IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 23, no. 11, pp. 1713-1726.
[18] J. Karvanen, J. Eriksson, and V. Koivunen, 2002. Adaptive Score Functions for Maximum Likelihood ICA, The Journal of VLSI Signal Processing, Vol. 32, no 1-2, PP 83-92.
[19] J. Karvanen, J. Eriksson, and V. Koivunen, June 2000. Source distribution adaptive maximum likelihood estimation of ICA model, in Proceedings of the 2nd International Conference on ICA and BSS, Helsinki, Finland, pp. 227– 232.
[20] A. N. Ebraheim, 2014. Exponentiated transmuted Weibull distribution a generalization of the Weibull distribution, International Journal of Mathematical, Computational, Physical and Quantum Engineering Vol. 8 No. 6.
[21] M. Li and J. Mao, June 2004. A new algorithm of evolutional blind source separation based on genetic algorithm, in Proceedings of the 5th World Congress on Intelligent Control and Automation, Hangzhou, Zhejiang, China, pp. 2240–2244.
[22] S. Mavaddaty and A. Ebrahimzadeh, December 2009. Evaluation of performance of genetic algorithm for speech signals separation, in Proceedings of the International Conference on Advances in Computing, Control and Telecommunication Technologies (ACT'09), Trivandrum, Kerala, India, pp. 681–683.
[23] A. Hyvarinen, J. Karhunen, and E. Oja, 2001. Independent Component Analsysis, JohnWiley & Sons.
[24] Internet web: http://sipi.usc.edu/database/database.cgi
[25] Internet web: http://www.lupusimages.com

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

IJCSIS REVIEWERS' LIST

CALL FOR PAPERS
International Journal of Computer Science and Information Security

IJCSIS 2016

ISSN: 1947-5500
http://sites.google.com/site/ijcsis/

International Journal Computer Science and Information Security, IJCSIS, is the premier scholarly venue in the areas of computer science and security issues. IJCSIS 2011 will provide a high profile, leading edge platform for researchers and engineers alike to publish state-of-the-art research in the respective fields of information technology and communication security. The journal will feature a diverse mixture of publication articles including core and applied computer science related topics.

Authors are solicited to contribute to the special issue by submitting articles that illustrate research results, projects, surveying works and industrial experiences that describe significant advances in the following areas, but are not limited to. Submissions may span a broad range of topics, e.g.:

Track A: Security

Access control, Anonymity, Audit and audit reduction & Authentication and authorization, Applied cryptography, Cryptanalysis, Digital Signatures, Biometric security, Boundary control devices, Certification and accreditation, Cross-layer design for security, Security & Network Management, Data and system integrity, Database security, Defensive information warfare, Denial of service protection, Intrusion Detection, Anti-malware, Distributed systems security, Electronic commerce, E-mail security, Spam, Phishing, E-mail fraud, Virus, worms, Trojan Protection, Grid security, Information hiding and watermarking & Information survivability, Insider threat protection, Integrity
Intellectual property protection, Internet/Intranet Security, Key management and key recovery, Language-based security, Mobile and wireless security, Mobile, Ad Hoc and Sensor Network Security, Monitoring and surveillance, Multimedia security ,Operating system security, Peer-to-peer security, Performance Evaluations of Protocols & Security Application, Privacy and data protection, Product evaluation criteria and compliance, Risk evaluation and security certification, Risk/vulnerability assessment, Security & Network Management, Security Models & protocols, Security threats & countermeasures (DDoS, MiM, Session Hijacking, Replay attack etc,), Trusted computing, Ubiquitous Computing Security, Virtualization security, VoIP security, Web 2.0 security, Submission Procedures, Active Defense Systems, Adaptive Defense Systems, Benchmark, Analysis and Evaluation of Security Systems, Distributed Access Control and Trust Management, Distributed Attack Systems and Mechanisms, Distributed Intrusion Detection/Prevention Systems, Denial-of-Service Attacks and Countermeasures, High Performance Security Systems, Identity Management and Authentication, Implementation, Deployment and Management of Security Systems, Intelligent Defense Systems, Internet and Network Forensics, Large-scale Attacks and Defense, RFID Security and Privacy, Security Architectures in Distributed Network Systems, Security for Critical Infrastructures, Security for P2P systems and Grid Systems, Security in E-Commerce, Security and Privacy in Wireless Networks, Secure Mobile Agents and Mobile Code, Security Protocols, Security Simulation and Tools, Security Theory and Tools, Standards and Assurance Methods, Trusted Computing, Viruses, Worms, and Other Malicious Code, World Wide Web Security, Novel and emerging secure architecture, Study of attack strategies, attack modeling, Case studies and analysis of actual attacks, Continuity of Operations during an attack, Key management, Trust management, Intrusion detection techniques, Intrusion response, alarm management, and correlation analysis, Study of tradeoffs between security and system performance, Intrusion tolerance systems, Secure protocols, Security in wireless networks (e.g. mesh networks, sensor networks, etc.), Cryptography and Secure Communications, Computer Forensics, Recovery and Healing, Security Visualization, Formal Methods in Security, Principles for Designing a Secure Computing System, Autonomic Security, Internet Security, Security in Health Care Systems, Security Solutions Using Reconfigurable Computing, Adaptive and Intelligent Defense Systems, Authentication and Access control, Denial of service attacks and countermeasures, Identity, Route and

Location Anonymity schemes, Intrusion detection and prevention techniques, Cryptography, encryption algorithms and Key management schemes, Secure routing schemes, Secure neighbor discovery and localization, Trust establishment and maintenance, Confidentiality and data integrity, Security architectures, deployments and solutions, Emerging threats to cloud-based services, Security model for new services, Cloud-aware web service security, Information hiding in Cloud Computing, Securing distributed data storage in cloud, Security, privacy and trust in mobile computing systems and applications, **Middleware security & Security features:** middleware software is an asset on
its own and has to be protected, interaction between security-specific and other middleware features, e.g., context-awareness, **Middleware-level security monitoring and measurement:** metrics and mechanisms for quantification and evaluation of security enforced by the middleware, **Security co-design:** trade-off and co-design between application-based and middleware-based security, **Policy-based management:** innovative support for policy-based definition and enforcement of security concerns, **Identification and authentication mechanisms:** Means to capture application specific constraints in defining and enforcing access control rules, **Middleware-oriented security patterns:** identification of patterns for sound, reusable security, **Security in aspect-based middleware:** mechanisms for isolating and enforcing security aspects, **Security in agent-based platforms:** protection for mobile code and platforms, Smart Devices: Biometrics, National ID cards, Embedded Systems Security and TPMs, RFID Systems Security, Smart Card Security, Pervasive Systems: Digital Rights Management (DRM) in pervasive environments, Intrusion Detection and Information Filtering, Localization Systems Security (Tracking of People and Goods), Mobile Commerce Security, Privacy Enhancing Technologies, Security Protocols (for Identification and Authentication, Confidentiality and Privacy, and Integrity), Ubiquitous Networks: Ad Hoc Networks Security, Delay-Tolerant Network Security, Domestic Network Security, Peer-to-Peer Networks Security, Security Issues in Mobile and Ubiquitous Networks, Security of GSM/GPRS/UMTS Systems, Sensor Networks Security, Vehicular Network Security, Wireless Communication Security: Bluetooth, NFC, WiFi, WiMAX, WiMedia, others

This Track will emphasize the design, implementation, management and applications of computer communications, networks and services. Topics of mostly theoretical nature are also welcome, provided there is clear practical potential in applying the results of such work.

Track B: Computer Science

Broadband wireless technologies: LTE, WiMAX, WiRAN, HSDPA, HSUPA, Resource allocation and interference management, Quality of service and scheduling methods, Capacity planning and dimensioning, Cross-layer design and Physical layer based issue, Interworking architecture and interoperability, Relay assisted and cooperative communications, Location and provisioning and mobility management, Call admission and flow/congestion control, Performance optimization, Channel capacity modeling and analysis, Middleware Issues: Event-based, publish/subscribe, and message-oriented middleware, Reconfigurable, adaptable, and reflective middleware approaches, Middleware solutions for reliability, fault tolerance, and quality-of-service, Scalability of middleware, Context-aware middleware, Autonomic and self-managing middleware, Evaluation techniques for middleware solutions, Formal methods and tools for designing, verifying, and evaluating, middleware, Software engineering techniques for middleware, Service oriented middleware, Agent-based middleware, Security middleware, Network Applications: Network-based automation, Cloud applications, Ubiquitous and pervasive applications, Collaborative applications, RFID and sensor network applications, Mobile applications, Smart home applications, Infrastructure monitoring and control applications, Remote health monitoring, GPS and location-based applications, Networked vehicles applications, Alert applications, Embeded Computer System, Advanced Control Systems, and Intelligent Control : Advanced control and measurement, computer and microprocessor-based control, signal processing, estimation and identification techniques, application specific IC's, nonlinear and adaptive control, optimal and robot control, intelligent control, evolutionary computing, and intelligent systems, instrumentation subject to critical conditions, automotive, marine and aero-space control and all other control applications, Intelligent Control System, Wiring/Wireless Sensor, Signal Control System. Sensors, Actuators and Systems Integration : Intelligent sensors and actuators, multisensor fusion, sensor array and multi-channel processing, micro/nano technology, microsensors and microactuators, instrumentation electronics, MEMS and system integration, wireless sensor, Network Sensor, Hybrid

Sensor, Distributed Sensor Networks. Signal and Image Processing : Digital signal processing theory, methods, DSP implementation, speech processing, image and multidimensional signal processing, Image analysis and processing, Image and Multimedia applications, Real-time multimedia signal processing, Computer vision, Emerging signal processing areas, Remote Sensing, Signal processing in education. Industrial Informatics: Industrial applications of neural networks, fuzzy algorithms, Neuro-Fuzzy application, bioInformatics, real-time computer control, real-time information systems, human-machine interfaces, CAD/CAM/CAT/CIM, virtual reality, industrial communications, flexible manufacturing systems, industrial automated process, Data Storage Management, Harddisk control, Supply Chain Management, Logistics applications, Power plant automation, Drives automation. Information Technology, Management of Information System : Management information systems, Information Management, Nursing information management, Information System, Information Technology and their application, Data retrieval, Data Base Management, Decision analysis methods, Information processing, Operations research, E-Business, E-Commerce, E-Government, Computer Business, Security and risk management, Medical imaging, Biotechnology, Bio-Medicine, Computer-based information systems in health care, Changing Access to Patient Information, Healthcare Management Information Technology. Communication/Computer Network, Transportation Application : On-board diagnostics, Active safety systems, Communication systems, Wireless technology, Communication application, Navigation and Guidance, Vision-based applications, Speech interface, Sensor fusion, Networking theory and technologies, Transportation information, Autonomous vehicle, Vehicle application of affective computing, Advance Computing technology and their application : Broadband and intelligent networks, Data Mining, Data fusion, Computational intelligence, Information and data security, Information indexing and retrieval, Information processing, Information systems and applications, Internet applications and performances, Knowledge based systems, Knowledge management, Software Engineering, Decision making, Mobile networks and services, Network management and services, Neural Network, Fuzzy logics, Neuro-Fuzzy, Expert approaches, Innovation Technology and Management : Innovation and product development, Emerging advances in business and its applications, Creativity in Internet management and retailing, B2B and B2C management, Electronic transceiver device for Retail Marketing Industries, Facilities planning and management, Innovative pervasive computing applications, Programming paradigms for pervasive systems, Software evolution and maintenance in pervasive systems, Middleware services and agent technologies, Adaptive, autonomic and context-aware computing, Mobile/Wireless computing systems and services in pervasive computing, Energy-efficient and green pervasive computing, Communication architectures for pervasive computing, Ad hoc networks for pervasive communications, Pervasive opportunistic communications and applications, Enabling technologies for pervasive systems (e.g., wireless BAN, PAN), Positioning and tracking technologies, Sensors and RFID in pervasive systems, Multimodal sensing and context for pervasive applications, Pervasive sensing, perception and semantic interpretation, Smart devices and intelligent environments, Trust, security and privacy issues in pervasive systems, User interfaces and interaction models, Virtual immersive communications, Wearable computers, Standards and interfaces for pervasive computing environments, Social and economic models for pervasive systems, Active and Programmable Networks, Ad Hoc & Sensor Network, Congestion and/or Flow Control, Content Distribution, Grid Networking, High-speed Network Architectures, Internet Services and Applications, Optical Networks, Mobile and Wireless Networks, Network Modeling and Simulation, Multicast, Multimedia Communications, Network Control and Management, Network Protocols, Network Performance, Network Measurement, Peer to Peer and Overlay Networks, Quality of Service and Quality of Experience, Ubiquitous Networks, Crosscutting Themes – Internet Technologies, Infrastructure, Services and Applications; Open Source Tools, Open Models and Architectures; Security, Privacy and Trust; Navigation Systems, Location Based Services; Social Networks and Online Communities; ICT Convergence, Digital Economy and Digital Divide, Neural Networks, Pattern Recognition, Computer Vision, Advanced Computing Architectures and New Programming Models, Visualization and Virtual Reality as Applied to Computational Science, Computer Architecture and Embedded Systems, Technology in Education, Theoretical Computer Science, Computing Ethics, Computing Practices & Applications

Authors are invited to submit papers through e-mail ijcsiseditor@gmail.com. Submissions must be original and should not have been published previously or be under consideration for publication while being evaluated by IJCSIS. Before submission authors should carefully read over the journal's Author Guidelines, which are located at http://sites.google.com/site/ijcsis/authors-notes .

ISSN 1947 5500
http://sites.google.com/site/ijcsis/