IJACSA
WHERE WISDOM SHARES
International Journal of Advanced Computer Science and Applications

Volume 3 Issue 3

March 2012

ISSN 2156-5570(Online)
ISSN 2158-107X(Print)

SAI

www.ijacsa.thesai.org

INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS

THE SCIENCE AND INFORMATION ORGANIZATION

www.thesai.org | info@thesai.org

Editorial Preface

From the Desk of Managing Editor…

IJACSA seems to have a cult following and was a humungous success during 2011. We at The Science and Information Organization are pleased to present the March 2012 Issue of IJACSA.

While it took the radio 38 years and the television a short 13 years, it took the World Wide Web only 4 years to reach 50 million users. This shows the richness of the pace at which the computer science moves. As 2012 progresses, we seem to be set for the rapid and intricate ramifications of new technology advancements.

With this issue we wish to reach out to a much larger number with an expectation that more and more researchers get interested in our mission of sharing wisdom. The Organization is committed to introduce to the research audience exactly what they are looking for and that is unique and novel. Guided by this mission, we continuously look for ways to collaborate with other educational institutions worldwide.

Well, as Steve Jobs once said, Innovation has nothing to do with how many R&D dollars you have, it's about the people you have. At IJACSA we believe in spreading the subject knowledge with effectiveness in all classes of audience. Nevertheless, the promise of increased engagement requires that we consider how this might be accomplished, delivering up-to-date and authoritative coverage of advanced computer science and applications.

Throughout our archives, new ideas and technologies have been welcomed, carefully critiqued, and discarded or accepted by qualified reviewers and associate editors. Our efforts to improve the quality of the articles published and expand their reach to the interested audience will continue, and these efforts will require critical minds and careful consideration to assess the quality, relevance, and readability of individual articles.

To summarise, the journal has offered its readership thought provoking theoretical, philosophical, and empirical ideas from some of the finest minds worldwide. We thank all our readers for their continued support and goodwill for IJACSA. We will keep you posted on updates about the new programmes launched in collaboration.

We would like to remind you that the success of our journal depends directly on the number of quality articles submitted for review. Accordingly, we would like to request your participation by submitting quality manuscripts for review and encouraging your colleagues to submit quality manuscripts for review. One of the great benefits we can provide to our prospective authors is the mentoring nature of our review process. IJACSA provides authors with high quality, helpful reviews that are shaped to assist authors in improving their manuscripts.

We regularly conduct surveys and receive extensive feedback which we take very seriously. We beseech valuable suggestions of all our readers for improving our publication.

Thank you for Sharing Wisdom!

Managing Editor
IJACSA
Volume 3 Issue 3 March 2012
ISSN 2156-5570 (Online)
ISSN 2158-107X (Print)

Editorial Board

Reviewer Board Members

CONTENTS

Paper 1: A New Approach for Arabic Handwritten Postal Addresses Recognition
Authors: Moncef Charfi, Monji Kherallah, Abdelkarim El Baati, Adel M. Alimi
PAGE 1 – 7

Paper 2: A Keyword Driven Framework for Testing Web Applications
Authors: Rashmi, Neha Bajpai
PAGE 8 – 14

Paper 3: Effect of Error Packetization on the Quality of Streaming Video in Wireless Broadband Networks
Authors: Aderemi A. Atayero, Oleg I. Sheluhin, Yury A. Ivanov
PAGE 15 – 19

Paper 4: An Overview of Video Allocation Algorithms for Flash-based SSD Storage Systems
Authors: Jaafer Al-Sabateen, Saleh Ali Alomari, Putra Sumari
PAGE 20 – 25

Paper 5: Building Trust In Cloud Using Public Key Infrastructure
Authors: Ms. Heena Kharche, Mr. Deepak Singh Chouhan
PAGE 26 – 31

Paper 6: Contextual Modelling of Collaboration System
Authors: Wafaa DACHRY, Brahim AGHEZZAF, Bahloul BENSASSI, Adil SAYOUTI
PAGE 32 – 37

Paper 7: Development of a Mobile Phone Based e-Health Monitoring Application
Authors: Duck Hee Lee, Ahmed Rabbi, Jaesoon Choi, Reza Fazel-Rezai
PAGE 38 – 43

Paper 8: Development of knowledge Base Expert System for Natural treatment of Diabetes disease
Authors: Sanjeev Kumar Jha, D.K.Singh
PAGE 44 – 47

Paper 9: Maximum-Bandwidth Node-Disjoint Paths
Authors: Mostafa H. Dahshan
PAGE 48 – 56

Paper 10: Message Segmentation to Enhance the Security of LSB Image Steganography
Authors: Dr. Mohammed Abbas Fadhil Al-Husainy
PAGE 57 – 62

Paper 11: Mobile Learning Environment System (MLES): The Case of Android-based Learning Application on Undergraduates' Learning
Authors: Hafizul Fahri Hanafi, Khairulanuar Samsudin
PAGE 63 – 66

Paper 12: Simple and Efficient Contract Signing Protocol
Authors: Abdullah M. Alaraj
PAGE 67 – 71

Paper 13: The Use of Information and Communication Technologies (ICT) in Front Office Operations of Chain Hotels in Ghana
Authors: Albert Kwansah Ansah, Victoria S. Blankson, Millicent Kontoh
PAGE 72 – 77

Paper 14: A Digital Ecosystem-based Framework for Math Search Systems
Authors: Mohammed Q. Shatnawi, Qusai Q. Abuein
PAGE 78 – 83

Paper 15: OFW-ITS-LSSVM: Weighted Classification by LS-SVM for Diabetes diagnosis
Authors: Fawzi Elias Bekri, Dr. A. Govardhan
PAGE 84 – 93

Paper 16: OCC: Ordered congestion control with cross layer support in Manet routing
Authors: T.Suryaprakash Reddy, Dr.P.Chenna Reddy
PAGE 94 – 101

Paper 17: Multi-Objective Intelligent Manufacturing System for Multi Machine Scheduling
Authors: Sunita Bansal, Dr. Manuj Darbari
PAGE 102 – 105

Paper 18: Evaluation of Data Security Measures in a Network Environment Towards Developing Cooperate Data Security Guidelines
Authors: Ayub Hussein Shirandula
PAGE 106 – 109

Paper 19: RC4 stream cipher and possible attacks on WEP
Authors: Lazar Stošić, Milena Bogdanović
PAGE 110 – 114

Paper 20: Transforming Conceptual Model into Logical Model for Temporal Data Warehouse Security: A Case Study
Authors: Marwa S.Farhan, Mohamed E. Marie, Laila M. El-Fangary, Yehia K. Helmy
PAGE 115 – 123

Paper 21: Web Anomaly Misuse Intrusion Detection Framework for SQL Injection Detection
Authors: Shaimaa Ezzat Salama, Mohamed I. Marie, Laila M. El-Fangary, Yehia K. Helmy
PAGE 124 – 129

Paper 22: An Improved Grunwald-Letnikov Fractional Differential Mask for Image Texture Enhancement
Authors: Vishwadeep Garg, Kulbir Singh
PAGE 130 – 135

Paper 23: A Comparative Study on Temporal Mobile Access Pattern Mining Methods
Authors: Hanan Fahmy, Maha A.Hana, Yahia K. Helmy
PAGE 136 – 139

Paper 24: A Schema for Generating Update Semantics
Authors: Jos´e Luis Carballido Carranza, Claudia Zepeda, Guillermo Flores
PAGE 140 – 147

A New Approach for Arabic Handwritten Postal Addresses Recognition

Moncef Charfi, Monji Kherallah, Abdelkarim El Baati, Adel M. Alimi
REGIM: Research Group on Intelligent Machines,
National Engineering School of Sfax, University of Sfax, Tunisia

***Abstract*—In this paper, we propose an automatic analysis system for the Arabic handwriting postal addresses recognition, by using the beta elliptical model. Our system is divided into different steps: analysis, pre-processing and classification. The first operation is the filtering of image. In the second, we remove the border print, stamps and graphics. After locating the address on the envelope, the address segmentation allows the extraction of postal code and city name separately. The pre-processing system and the modeling approach are based on two basic steps. The first step is the extraction of the temporal order in the image of the handwritten trajectory. The second step is based on the use of Beta-Elliptical model for the representation of handwritten script. The recognition system is based on Graph-matching algorithm. Our modeling and recognition approaches were validated by using the postal code and city names extracted from the Tunisian postal envelopes data. The recognition rate obtained is about 98%.**

Keywords-Postal automation; handwritten postal address; address segmentation; beta-elliptical representation; graph matching.

I. INTRODUCTION

For several years, on-line and off-line hand-writing recognition has been considered [8], [24]. The postal automation, bank checks identification, automatic processing of administrative files and cultural patrimony heritage are direct applications of the optical character recognition, which present difficult problems due to the presence of handwritten manuscripts in such documents.

Postal automation has constituted a potential application of character recognition and a real driving challenge feeding research in such domain. Everywhere in the industrialized world, the postal services have financed and continue to finance lots of works in such complex domain. Today, in the western countries, millions of postal objects essentially made up of letters and parcels, are collected in mail sorting centres for their redistribution. These systems recognize the postal address, and print a bar code on the envelope. Letter forwarding is based on the reading of the addresses' bar code, thus making it possible to analyze 50000 letters per hour [32],[35].

On the other hand, the Arab and Eastern countries are somewhat behind in this domain. To take up the challenge, we have considered the problem of the automatic process of Arabic postal addresses in Tunisia and Arab countries.

Our paper is formulated as follows: section 2 presents a state of the art of handwritten character recognition, and the problems encountered in the process of Arabic postal addresses. Paragraph 3, describes the pre-processing steps, and paragraph 4 gives the temporal order reconstruction. In section 5 we present the beta elliptical approach for handwriting modelling. Section 6 is devoted to the recognition and some experimental works discussion.

II. STATE OF THE ART AND PROBLEMATIC

The postal automation systems are generally based on printed or off line handwriting recognition. Printed characters are now well recognized, but handwritten character recognition remains very difficult, on account of its very great variability. We find in the literature several descriptions of such systems. Wada [33] proposed a total system of automatic analysis of addresses. Gilloux presented a postal addresses recognition system [13]. Heute proposed a system of postal automation as a potential application for recognition digits and characters manuscripts [15], [16]. In such systems, the envelopes submitted to recognition are ordinary (handwritten or printed) and present the greater part of the difficulties that a recognition system has to face: address location, separation printed/manuscript, segmentation in lines, words and characters, inclination correction, etc. In [3], text/graphic segmentation which is based on the detection of the geometric model detection of the related components on an envelope, and on graph discrimination such as stamps and logos, is done by computing pixel density. Menoti and all, present a segmentation algorithm based on feature selection in wavelet space. The aim is to automatically separate in postal envelopes the regions related to background, stamps, rubber stamps, and the address blocks [26], [34], [36].

The techniques developed for OCR systems are generally based on the neural networks and Markovian approaches [31]. In fact, several recognition systems in the literature are based on Hidden Markov Models [20], [37]. Lee and all [21] have developed a new hybrid approach to the verification of handwritten addresses in Singapore. The hybrid verification system seeks to reduce the error rate by the correlation of the extracted postcode features set recognized words from the original handwritten address. Novel use of syntactic features extracted from words has resulted in a significant reduction in the error rate while keeping the recognition rate high [30].

For the degraded character recognition, Likforman-Sulem and Sigelle proposed in 2008 models based on the formalism of Bayesian networks to re-introduce the notion of spatial context [22]. Bayesian network is a probabilistic graphical model in which links are introduced between distant observations. Similarly, models based on recurrent neural networks extend classical neural networks by introducing the notion of context by bidirectional dependencies. In particular the model LSTM (Long Short Term Memory) has been recently applied to the recognition of cursive online [14].

Gaceb et all, present a new approach for address block location based on pyramidal data organization and on a hierarchical graph colouring for classification process. This new approach permits to guarantee a good coherence between different modules and to reduce the computation time and the rejection rate, and gives satisfying rate of 98% of good location [11], [12].

Liu et all, proposes an approach to retrieve envelope images from a large image database, by graph matching. The attributes of nodes and edges in the graph are described by characteristics of the envelope image, and the results of experiments tests using this approach are promising [23], [29].

In Arabic optical character recognition (AOCR), Lorigo presented a study focusing on the off-line Arabic handwriting recognition. She proved that Arabic writing presents important technical challenges which have to take up [2], [17], [24], [28] and [31].

In the case of on-line handwriting, the dynamic information, such as the temporal order and pen speed, make a performance for handwriting modelling. In this context, the reconstruction of the temporal order of off-line handwriting improves the performance of recognition systems [1] [7].

In the case of the printed and scanned documents, the results are very interesting if the images are high quality. However, these results decrease quickly if the image contains noise as stain, background, etc. Based on the complementarities existing between classifiers (Multi-Layer Perceptron (MLP) [5], Hidden Markov Model (HMM) [10], Genetic Algorithm (GA) [19], Fuzzy Logic [6], etc.),

The majority of classifiers meet a major problem which lies in the variability of the vector features size. In literature, three approaches are commonly used to manage the problems of dimensionality. These approaches are: Genetic Algorithms, Dynamic Programmation and Graph Matching. In our work, the method of graph matching is appropriately used to cover the problem of dimensionality of feature vector, as was done by Namboodiri and Jain in the case of Graph Matching use for Arabic word recognition, and by Rokbani and al., in the case of the Genetic Algorithm and Graph Matching combination for on-line Arabic word recognition [25], [28].

The researcher in the Arabic postal envelopes processing encounters many difficulties: extraction of addressee address, extraction of postal code, segmentation of postal code in digits and variability of writer.

In the absence of standard format of envelopes, locating the address becomes difficult and depends on locating all objects constituting the address [4], [33], [35]: analysis, location and description of interest zone. Besides, it's noticeable that the handwritten address presents some difficulties and problems, like the inclination of the baseline of the handwritten address, and the variable position of the address on the envelope, which requires correction and particular processing.

Addressee's address, deriving from the extraction, contains the numerical and alphabetic data. A stage of discrimination (digit/letter) is therefore necessary. The position of the postal code in the address is variable according to the writer.

Analysis of the postal code depends on its decomposition into digits. But, this decomposition is not always possible because several types of connections exist between the handwritten digits:

- Simple connections, where only one link exists between two strokes of writing;
- Multiple connections, where at least three strokes of writings exist on at least one connection [15], [24].

The pattern of handwritten writing is very variable. It translates the style of writing, the mood and the writer's personality, which makes it difficult to characterize (see figure 1):

- The chosen value of image resolution is important because it conditions the recognition system. It is very important to choose a high resolution (≥ 300 ppi) to keep the maximum of information on the image.
- Noise and irregularities in the tracing are at the origin of the disconnection of some features.
- The distortions are the dilations, narrowing and other local variations in the writing.
- The variability of the style, that is the use of different forms to represent the same character, such as the flowery style or the slope and, in a general manner, everything that characterizes the writer's habits.
- The translation of the whole character or a part of its components.
- The dissymmetry that leads to a closure and a complete closing of the digit.
- The bad adjusting of the digits causing abrupt interruptions of the lines.

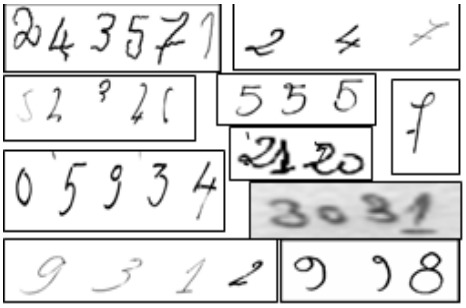


Figure 1. General problems in digits recognition

- The slope of the digits and the baseline: different directions can be seen on the same page because the changes of paper's orientations during writing.
- A handwritten digit cannot represent the same features as the set of other digits of same family, and therefore generates confusions [14] (see figure 1).

III. Pre-Processing Steps

The pre-processing steps consists on: filtering of the image, location and suppression of printed borders, location and suppression of stamps and graphics, address location, segmentation of address in lines, segmentation of the last two lines, and segmentation into connected components [5].

Figure 2. Pre-processing steps

(a) Image of envelope,
(b) Location and suppression of printed borders,
(c) Suppression of stamps and graphics,
(d) Address location,
(e) Address segmentation in lines by horizontal projection and in words by vertical projection,
(f) Words of address segmented.

In figure 2, we regrouped all illustrations of the main pre-processing steps.

The final step is the discrimination postal code/city name that is based on the test of regularity of connected components and test of eccentricity.

IV. Temporal Order Reconstruction

The handwritten of a word is the stroke constituted by a set of curved lines. Every line has a starting point and an ending point. Before the temporal order reconstruction, we firstly proceed to skeleton segmentation. Word image, representing the city name, goes through three stages of pre-processing: binarization, filtering, skeletisation and elimination of the diacritic signs, as the points above and below the words and the vowel signs, (see figure 3). Three types of characteristic points will be extracted from the skeleton of the tracing [7]:

- The end stroke point: this is the black pixel that possesses only one neighbour of the same type.
- The branching point: this is the black pixel that possesses three neighbours of the same type.
- The crossing point: this is the black pixel that possesses four neighbours of the same type.

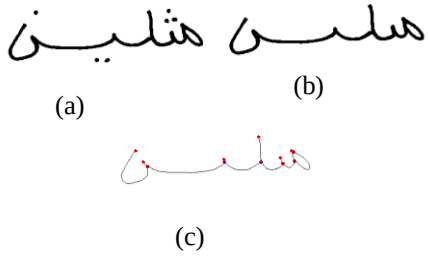


Figure 3. Pre-processing steps

(a) Original image of Tunisian city name "Methline"
(b) Suppression of diacritics
(c) The same word after the pre-processing steps and characteristic point's detection

An algorithm of segmentation of the skeleton makes the segments of a word. These segments are classified into three categories: (see figure 4)

- Segment 1: represents a stroke that is located between two end points or between an end point and a branching point (or crossing point).
- Segment 2: shows a stroke of link that is located between two branching points (or crossing) or between a branching point and a crossing point. This type of segment does not represent a contour of an occlusion.
- Segment 0: presents a stroke of link that is located between two branching points (or crossing) or between a branching point and a crossing point but it represents a contour of an occlusion.

The segmentation is released by an inspection of successive skeleton points. In an instance, one segment is limited between two characteristic points. As a result, this segmentation allows a first organization of the points of every segment. To facilitate this operation, we start with the extraction of the segments of type 1, type 2 and type 0. We eliminate the segments localized on the skeleton every time.

The temporal order reconstruction is made by the consideration of these criteria:

- Choose the direction right-left of the pixels displacement.
- Choose the minimum distance between pixels.
- Minimize the repetition of the segments.

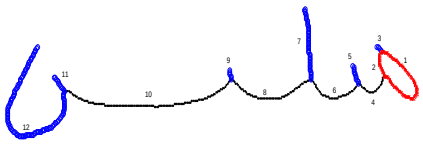


Figure 4. Segmentation of the skeleton of an Arabic word.

- Choose the minimum angular deviation in the crossing and branching points [27].

The sequence of these segments represents the original trajectory of the handwritten word. The rebuilt signal will be represented by a succession of X and Y coordinates.

The rebuilt signal does not contain the speed of the pen. A study made in the neuromuscular effect shows that the pen speed decreases at the beginning and the end of the stroke and in the angular variation of the curve. That is remarkable, if we observe an on-line signal acquired by a tablet. The on-line signal is acquired in real time. If we take into account the resolution (or the time of recording) of the tablet, it is noticed clearly that these points are not distributed in an equidistant manner. Concentrations of points are observed at the beginning and the end of the feature and in the curves of the stroke. This information is used in the on- line systems to calculate curvilinear speed. In order to obtain this information and to profit from on-line modelling of the writing, the rebuilt signal will be sampled by applying a method described in [9]. In order to modelling the rebuilt signal, we use the beta-elliptic representation developed by [18].

V. Beta- Elliptic Approach For Handwriting Modelling

As it was explained in [17], the Beta-elliptic model considers a simple movement as the response to the neuromuscular system, which is described by an elliptic trajectory and a Beta velocity profile. Handwritten scripts are, then, segmented into simple movements, as already mentioned, called strokes, and are the result of a superimposition of time-overlapped velocity profiles. In our approach of modelling, a simple stroke is approximated by a Beta profile in the dynamic domain which corresponds in turn to an elliptic arc in the static domain. As it was explained in [17], [18], the complete velocity profile of the neuromuscular system will be described

$$\beta(t, q, p, t_0, t_1)=\begin{cases}\left(\frac{t-t_0}{t_c-t_0}\right)^p\left(\frac{t_1-t}{t_1-t_c}\right)^q & \text{if } t \in [t_0, t_1] \\ 0 & elsewhere\end{cases} \quad (1)$$

by a Beta model as follows:

Where :

p and q are intermediate parameters, which have an influence on the symmetry and the width of Beta shape, t0 is the starting time of Beta function, tc is the instant when the curvilinear velocity reaches the amplitude of the inflexion point, t1 is the ending time of Beta function, $t0 < t1 \in IR$ (IR is the real set) and :

$$t_c = \frac{p \times t_1 + q \times t_0}{p+q} \quad (2)$$

We also consider the geometric representation of the handwritten trajectory. In the geometric plan, the trajectory is represented by a sequence of elliptic arcs [17]. The elliptic model is a static model.

The elliptic equation is written as follows:

$$\frac{X^2}{a^2}+\frac{Y^2}{b^2}=1 \quad (3)$$

Consequently, a stroke is characterized by seven parameters. The first four Beta parameters (t0, t1, p and k) reflect the global timing properties of the neuromuscular networks involved in generating the movement (k=1), whereas the last three elliptic parameters (θ, a and b) describe the global geometric properties of the set of muscles and joints recruited to execute the movement, θ is the angle of elliptical stroke, a and b are respectively the big and small axes of the ellipse. The result of Beta model reconstruction of signal velocity is shown in figure 5.

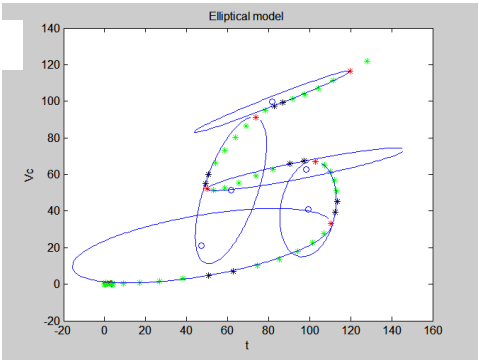


Figure 5. Example of elliptic representation of digit 5

VI. Recognition System And Experimental Results

The graph matching algorithm performs a distance computation between two trajectories. A nearest neighbourhood algorithm is used to associate the nearest points between graph trajectories. The Euclidian distance is then calculated in order to evaluate the graphic similarity. Note, that no deformation is assumed to the graphs during the processing. Figure 6 shows a fragment of handwritten graphs that had been superposed and scaled, then associated points (see figures 6a and 6b). If N1 and N2 are respectively the number of strokes of graph 1 and graph 2, so the distance between the traces of strokes is calculated by the formula 4. Every stroke will be represented by a middle point [28].

$$Dist_{1,2} = \frac{1}{N}\sum_{i=1}^{N}(dist_{1,2})_i + P*(|N_1 - N_2|) \quad (4)$$

$$P = Max_{i=0}^{N}(dist_{1,2})_i \quad (5)$$

Where N= min (N_1, N_2), P is a penalty (formula 5)

and $(dist_{1,2})_i$ is the Euclidian distance between two associated points (formula 6).

$$(dist_{1,2})_i = \sqrt{(X_i^1 - X_i^2)^T * (X_i^1 - X_i^2)} \qquad (6)$$

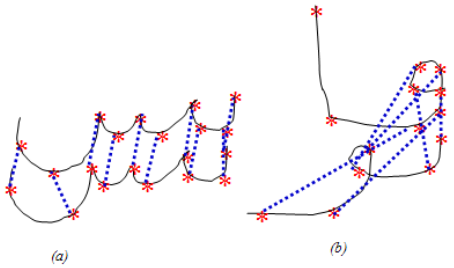


Figure 6. Graph matching process

(a) two similar graphs
(b) two different graphs

To test the developed system, we used the dataset of envelopes which have been created in our laboratory. It is made up of one thousand images of handwritten addresses, 740 of which are in Arabic and 260 in Latin symbols.

The addresses images are obtained via scanning envelopes collected from the education services files of two academic establishments in Sfax (see figure 7).

Figure 7. Some samples of images of addresses.

The scanning was done with 300 ppi resolution and 256 colours. The Tunisian postal code consists of 4 digits. Our data set of addresses consists of 1000 addresses enabling to have a data set of 4000 handwritten digits. The recognition system of the developed postal code address developed is divided into pre-processing steps and subsequent classifier. The processing steps were done by a serial mode which consists in filtering, smoothing, order reconstruction of the handwritten trajectory and the beta elliptic modelling representation.

The results are satisfactory and encouraging: 97% of the total envelopes images have been filtered and smoothed. 3% will be manually classified because of the existence of noises and imperfections in handwritten addresses.

In order to benefit of the on line feature, we reconstructed the temporal order of the handwritten words. Compared [1] to other systems, our modelling system based on neuro-physiological approach performs better. In fact, we reduced the vector size representing word by using the beta elliptic representation. To recognize the postal city name, we developed the graph matching algorithm. The recognition rate obtained is about 98%. This means that out of 100 envelopes, 98 are recognized and only 2 are released and will be processed manually.

VII. Conclusion

The postal automation is one of the most active OCR applications. It becomes indispensable to ensure a fast postal service. These systems have drawn more and more interest. Compared to results obtained using neural network [5], our results are more promising. In face of the complexity and the variability of the handwritten words, the results obtained are acceptable and very promising.

VIII. Future Works

We purport to pursue our work to improve our system related to processing the handwritten digits and characters. We hope to reach the performances of international systems conceived for the automatic processing of postal addresses. Also, we intend to extend our study to the areas of many administrative forms and bank checks processing applications.

Acknowledgments

This work is partly supported by the project financed by the "Fond International de Coopération Universitaire" (FICU). The authors would like to address particular thanks to M. Cheriet, G. Stamon, J. Suen, E. Lecolinet for their help and contribution during all the phases of project realization. Also, the authors would like to acknowledge the financial support of this work by grants from the General Department of Scientific Research and Technological Renovation (DGRST), Tunisia, under the ARUB program 01/UR/11/02.

References

[1] Abu Haiba I.S.I., Ahmed P., Restoration of temporal information in off-line Arabic handwriting. Pattern Recognition, Vol. 26, N° 7, pp.1009-1017, 1993.

[2] Aoki Y., Akagi T., Nakao A., Natori N. and Mizutani H., A new approach for multilingual address recognition. Proceeding of ICDAR'99, pp 313-316, 1999.

[3] Bennasri A., Zahour A., Taconet B., Méthode pour la séparation texte/graphique sur les adresses postales, 6ème Colloque International francophone de l'Écrit et du Document, CIFED'00, France 2000.

[4] Bochnia G. et Facon J., Segmentation du bloc adresse d'enveloppes postales complexes. Congrès International Francophone de l'Écrit et du Document, CIFED'2002, Hammamet, Tunisie, 22-25 Octobre 2002.

[5] Charfi M., Hamdani T.M., Alimi A. M., A New Decision-Making Intelligent System for the Recognition of Bilingual Handwritten Postal Addresses, Journal of Decision Systems, volume 14 - No. 1-2, pp 123-155, 2005.

[6] Charfi M., Alimi A. M., Recognition of printed Numerals with the Beta Fuzzy Neural Network, Proc. IEEE/IMACS Multiconference on Computational Engineering in Systems Applications: CESA'98, Hammamet, Tunisia, April 1998.

[7] Dreuw Ph., Rybach D., Gollan C. and Ney H., Writer Adaptive Training and Writing Variant Model Refinement for Offline Arabic Handwriting Recognition, Document Analysis and Recognition, International Conference on, pp. 21-25, 10th International Conference on Document Analysis and Recognition, Barcelona, Spain, 26-29 July 2009.

[8] Elabed H, Kherallah M, Märgner V, Alimi A.M, "Online Arabic Handwriting Recognition Competition", International Journal on Document Analysis and Recognition, Vol. 14, Num. 1, pp 15-23, 2010.

[9] Elbaati A, Kherallah M., Alimi A. M., Ennaji A., De Hors-Ligne Vers un Système de Reconnaissance En-Ligne: Application à la Modélisation de l'Écriture Arabe Manuscrite Ancienne. Proc. Int. Conf. ANAGRAM'2006. pp 30-37. Suisse, Fribourg, 2006.

[10] Fehri A.. Reconnaissance de textes arabes multi fonte à l'aide d'une approche hybride neuro-markoviennes. Thèse de doctorat. Université de Tunis II, 156 pages, Avril 1999.

[11] Gaceb D., Eglin V., Lebourgeois F., Emptoz H., Improvement of postal mail sorting system, IJDAR(11), No. 2, November 2008, pp. 67-80.

[12] Gaceb D., Eglin V., Lebourgeois F., Emptoz H., Robust Approach of Address Block Localization in Business Mail by Graph Coloring, International Arab Journal of Information Technology (IAJIT) 6(3):221-229, ISSN1683-3198. 2009.

[13] Gilloux M., La lecture automatique de documents à la poste: de la reconnaissance des adresses postales à la saisie des questionnaires du recensement de la population. Journée Thématique organisée par le GRCE, Novembre 1999.

[14] Graves A., Liwicki M., Fernandez S., Bertolami R., Bunke H., Schmidhuber J., A Novel Connectionist System for Improved Unconstrained Handwriting Recognition. IEEE PAMI, vol. 31, no. 5, 2009.

[15] Heutte L., Reconnaissance de Caractère Manuscrits: Application à la Lecture Automatique des Chèques et des Enveloppes Postales Thèse de doctorat, Université de Rouen 1994.

[16] Heutte L., Lecourtier Y. and Moreau J. V., Une Nouvelle Méthode d'Extraction de Projections et de Profils pour la Reconnaissance de Chiffres Manuscrits, Actes du CNED, Rouen, Juillet 1994.

[17] Jarousse C. et Viard-Gaudin C., Localisation du code postal par réseau de neurones sur bloc adresse manuscrit non contraint, 1er Colloque International Francophone sur l'Écrit et le Document CIFED'98, Université Laval, canada, 11-13 mai 1998.

[18] Kherallah M., Hadded L., Mitiche A., Alimi A. M., On-Line Recognition Of Handwritten Digits Based On Trajectory And Velocity Modelling. International Journal of Pattern Recognition Letter. Vol. 29. pp. 580-594. 2008.

[19] Kherallah M., Bouri F., And Alimi A. M., On-Line Arabic Handwriting Recognition System Based On Visual Encoding And Genetic Algorithm. Engineering Applications of Artificial Intelligence (2008), doi:10.1016/j.engappai.2008.05.010

[20] Koerich, A. Sabourin L. R., Suen C. Y., Recognition and verification of unconstrained handwritten words. IEEE Trans Pattern Anal Mach Intell, 27(10) pp. 1509-1522, Oct 2005.

[21] Lee C.K., Leedham C.G., A New Hybrid Approach to Handwritten Address Verification, IJCV(57), No. 2, May 2004, pp. 107-120.

[22] Likforman-Sulem L., Sigelle M., Recognition of degraded characters using Dynamic Bayesian Networks, Pattern Recognition, Vol. 41, pp. 3092-3103, 2008.

[23] Liu Li, Lu Y., Suen C. Y., Retrieval of Envelope Images Using Graph Matching, International Conference on Document Analysis and Recognition (ICDAR), Beijing, China, 18-21 Sept. 2011, pp 99-103.

[24] Lorigo L. M. and Govindaraju V., Offline Arabic handwriting recognition: a survey. IEEE Trans Pattern Anal Mach Intell, May 2006, 28(5) pp 712-724.

[25] Mahadevan U. and Srihari S. N., Parsing and recognition of city, state, and zip codes in handwritten addresses, proceeding of ICDAR'99, pp 325-328, 1999.

[26] Menoti D., Borges D. L., Facon J., Britto A. S., Segmentation of Postal Envelopes for Address Block Location: an approach based on feature selection in wavelet space, Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003), Edinburgh, Scotland 03-06 August 2003, 5 pages.

[27] Namboodiri A. M. and Jain A.K., Online handwritten script recognition. IEEE Trans Pattern Anal Mach Intell, Jan 2004, 26(1), pp 124-30.

[28] Plamondon R. and Privitera C. M., The segmentation of cursive handwriting: an approach based on off-line recovery of the motor-temporal information, IEEE Transactions on Image Processing, volume 8, No 1, pp. 80-91, 1999.

[29] Rokbani N., Kherallah M., and Alimi A. M., Global Recognition of Arabic words by Graph Matching, and Genetic algorithm. Tozeur, Tunisie. in Proc. ACIDCA-ICMI'2005, pp. 959-963.

[30] Sari T., Sellami M.. Proposition d'une approche hybride pour le tri postal multilingue. Actes du Huitième Colloque International Francophone sur l'Écrit et le Document CIFED'04, La Rochelle, 21 - 25 juin 2004, pp 297-302.

[31] Srihari S. Yang W. J. and Govindaraju V., Information theoric analysis of Postal address fields for automatic address interpretation, proceeding of ICDAR'99, pp 309-312, 1999.

[32] United States Postal Service, editor. United States Postal Service. Advanced Technology Conference, November 5-7, 1990.

[33] Wada M., Proposal for fully automated mail processing system for the 21st century, Pattern Recognition Letters 14, pp 281-290, 1993.

[34] Walischewski H., Learning regions of interest in postal automation, proceeding of ICDAR'99, pp 317-320, 1999.

[35] http://www.infres.enst.fr/~elc/GRCE

[36] Wichello P., Locating address blocs and postcodes in mail-piece images, ICPR'13, Vienna, 1996.

[37] Zheng Y., Doermann H. Li and D., A parallel-line detection algorithm based on HMM decoding. IEEE Trans Pattern Anal Mach Intell,, 27(5) pp 777-792, May 2005.

AUTHORS PROFILE

Moncef Charfi Profile

Moncef Charfi was born in Sfax (Tunisia), in 1950. He received his Ph.D. from the University Paris-Sud Orsay (France) in 1980. He is currently HDR in Computer Systems Engineering and Assistant Professor in the Department of Computer Engineering and Applied Mathematics at the National School of Engineers of Sfax University of Sfax,. Member in the Research Group on Intelligent Machine REGIM. He is interested in his research to the field of Document Analysis and Recognition and the processing of the old documents. He reviewed several scientific articles. He is IEEE member, and Arab Computer Society (ACS) member.

Monji Kherallah Profile

Monji Kherallah was born in Sfax, Tunisia, in 1963. He received the Engineer Diploma degree and the Ph.D both in electrical engineering, in 1989 and 2008, from University of Sfax (ENIS). For fourteen years ago, he was an engineer in Biotechnology Centre of Sfax. Now he teaching in Faculty of Science of Sfax and member in Research Group of Intelligent Machines: REGIM. His research interest includes the Handwritten Documents Analysis and Recognition. The techniques used are based on intelligent methods, such as neural network, fuzzy logic, genetic algorithm etc. He is one of the developers of the ADAB-Database (used by more than 50 research groups from more than 10 countries). He co-organized the Arabic Handwriting Recognition Competitions at the Online Arabic Handwriting Competitions at ICDAR 2009 and ICDAR 2011. He has more than 40 papers, including journal papers and book chapters. He is a member of IEEE and IEEE AESS Tunisia Chapter Chair, 2010 and 2011. He is reviewer of several international journals.

Abdelkarim El-Baati Profile

Abdelkarim ELBAATI was born in Chebba, Tunisia, in 1976. He received the Ph.D in electrical engineering, in 2009, from University of Sfax (ENIS). Now he teaching in Higher Institute of Applied Sciences and Technology of Mahdia and member in Research Group of Intelligent Machines: REGIM. His research interest includes the Handwritten Documents Analysis and Recognition. The techniques used are based on intelligent

methods, such as HMM, genetic algorithm etc. He has more than 10 papers, including journal papers. He is a member of IEEE . He is reviewer of PRL journal.

Adel M. Alimi Profile

Adel M. Alimi was born in Sfax (Tunisia) in 1966. He graduated in Electrical Engineering 1990, obtained a PhD and then an HDR both in Electrical & Computer Engineering in 1995 and 2000 respectively. He is now professor in Electrical & Computer Engineering at the University of Sfax. His research interest includes applications of intelligent methods (neural networks, fuzzy logic, evolutionary algorithms) to pattern recognition, robotic systems, vision systems, and industrial processes. He focuses his research on intelligent pattern recognition, learning, analysis and intelligent control of large scale complex systems. He is associate editor and member of the editorial board of many international scientific journals (e.g. "IEEE Trans. Fuzzy Systems", "Pattern Recognition Letters", "NeuroComputing", "Neural Processing Letters", "International Journal of Image and Graphics", "Neural Computing and Applications", "International Journal of Robotics and Automation", "International Journal of Systems Science", etc.).

He was guest editor of several special issues of international journals (e.g. Fuzzy Sets & Systems, Soft Computing, Journal of Decision Systems, Integrated Computer Aided Engineering, Systems Analysis Modelling and Simulations).

He is the Founder and Chair of many IEEE Chapter in Tunisia section, he is IEEE Sfax Subsection Chair (2011), IEEE ENIS Student Branch Counsellor (2011), IEEE Systems, Man, and Cybernetics Society Tunisia Chapter Chair (2011), IEEE Computer Society Tunisia Chapter Chair (2011), he is also Expert evaluator for the European Agency for Research. He was the general chairman of the International Conference on Machine Intelligence ACIDCA-ICMI'2005 & 2000. He is an IEEE senior member.

A Keyword Driven Framework for Testing Web Applications

[1]Rashmi
Centre for Development of Advanced Computing,
Noida, India.

[2]Neha Bajpai
Centre for Development of Advanced Computing,
Noida, India.

***Abstract*—The goal of this paper is to explore the use of Keyword driven testing for automated testing of web application. In Keyword driven testing, the functionality of the system-under-test is documented in a table as well as in step by- step instructions for each test. It involves the creation of modular, reusable test components. These components are then assembled into test scripts. These components can be parameterized to make them reusable across various test script. These test scripts can also be divided into various reusable actions. This saves a lot of recording procedure. The Existing tools for this testing uses Html, Xml, Spreadsheet, etc. to maintain the test steps. The test results are analyzed to create test reports.**

Keyword-driven testing; test automation; test script;, test results; Html; test reports; test result; recording.

I. INTRODUCTION

Testing is an integral part of the software development. The goal of software testing is to find faults from developed software and to make sure they get fixed. It is important to find the faults as early as possible because fixing them is more expensive in the later phases of the development. The purpose of testing is also to provide information about the current state of the developed software from the quality perspective.

On a high level, software testing can be divided into dynamic and static testing. The division to these two categories can be done based on whether the software is executed or not. Static testing means testing without executing the code. This can be done with different kinds of reviews. Reviewed items can be documents or code. Other static testing methods are static code analysis methods for example syntax correctness and code complexity analysis.

Dynamic testing is the opposite of static testing. The system under test is tested by executing it or parts of it. Dynamic testing can be divided to functional testing and non-functional testing

The purpose of functional testing is to verify that software corresponds to the requirements defined for the system. The focus on functional testing is to enter inputs to the system under test and verify the proper output and state. The non-functional testing means testing quality aspects of software. Benefits of non-functional testing are performance, security, usability, portability, reliability, and memory management testing.

Automation testing means execution of test cases in an automated way without manual intervention. It was originated from simply record and playback which makes engineers repeat the work. This can be achieved either by using a third party tool like RFT, QTP, etc or by developing an in-house tool suited to the testing need. Test automation includes various activities like test generation, reporting the test execution results, and test management. All these test automation activities can take place on all the different test levels. These test levels are unit testing, integration testing, system testing, and acceptance testing. [2]

Automating the testing is not an easy task. There are several issues that have to be taken into account. These issues are like unrealistic expectations, poor testing practice, and an expectation that automated tests will find a lot of new defects, a false sense of security, maintenance, technical problems, and organizational issues. [2]

The test automation frameworks have evolved over the time. They have evolved into three generations. [3] Figure 1 shows evolution of Test Automation. In the beginning, there was record and playback script creation. In this, there were only stand-alone test scripts. After this, comes the Functional Decomposition. It consists of reusable functional; test modules.

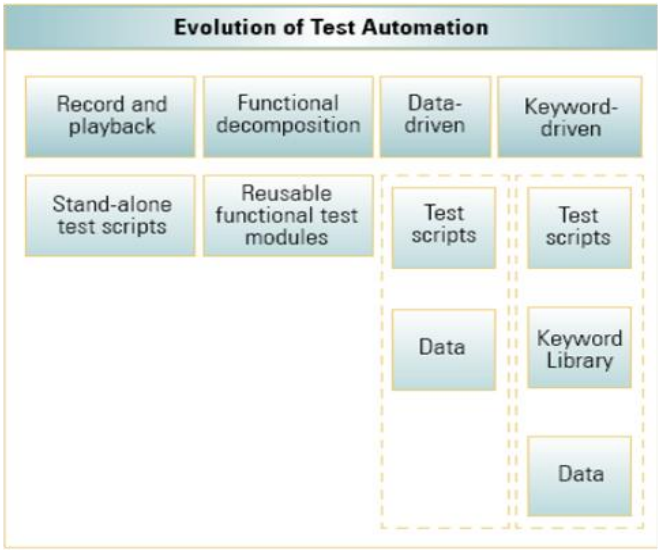


Figure 1. Evolution of Test Automation [7]

After that came data-driven testing. In this, test data is taken out of the scripts. This makes the test data variation easy and similar test cases can be created quickly.

Today, keyword-driven testing is getting more popular. It is a technique that separates much of the programming work from the actual test steps so that the test steps can be developed earlier and can be maintained with only minor updates. It consists of test scripts, keyword library and data. Table 1 shows Benefits and shortcomings of Automated Testing Approaches.

II. KEYWORD DRIVEN TESTING

It involves the creation of modular, reusable test components that are built by test architects and then assembled into test scripts by test designers. This removes the biggest limitation of the data-driven testing approach.

Keywords can be divided into base and user keywords. Base keywords are keywords implemented in the libraries. User keywords are keywords that are defined in the test data by combining base keywords or other user keywords [3]. The ability to create new user keywords in the test data de-creases the amount of needed base keywords and therefore amount of programming. The Test scripts can be added, deleted and modified. The test script modification helps in the parameterization of the test and in dividing a test into multiple actions.

TABLE I. BENEFITS AND SHORTCOMINGS OF AUTOMATED TESTING APPROACHES [7]

Approach	How it works	Benefits	Shortcomings
Record and Playback	Users' action are captured, then played back on the application	Ease of Scripting, not much technical expertise required	Difficult to maintain test scripts, not extendable, limited reusability, even small changes to the application require updates of scripts
Functional decomposition	Re-usable, repeatable snippet of functions are created	modular approach provides some flexibility, maintainability, reduces redundancy, larger test cases can be built in hierarchical fashion	Data exists within scripts, meaning limited reusability, ease of maintenance, depends largely on technical expertise, framework is high dependent on the framework.
Data-driven	Input/output data is maintained in external files	Size of the test pack is greatly reduced, improved maintainability	Depends on technical expertise of test team, maintenance and perpetuation are issues
Keyword-driven	Robust, application independent re-usable keyword libraries are built	Ease of maintenance and highly scalable reduced dependence on application availability	Requires great deal of efforts and is time consuming, expertise in test tool scripting language required by framework development

While testing a web application, there may be needed to check how the web application performs the same operations with multiple sets of data. For example, how a Web application responds to ten separate sets of data is to be checked.

Ten separate tests could be recorded, each with its own set of data. The tool can be test the application with this different data without the need of recording with these data. The test must be saved before running.

Actions divide the test into logical sections. When a new test is created, it contains a call to one action. By dividing the tests into calls to multiple actions, more modular and efficient tests can be designed. This is another feature of reusability of keywords that makes Keyword driven Testing more efficient and modular than the Data Driven Testing.

Various tools for test automation that support this technique are also developed in industry, such as Mercury's Quick Test Professional (QTP) and WinRunner, IBM's Rational Functional Tester (RFT) and Robot on functional testing, and LoadRunner from Mercury, SilkPerformer from Borland, Grinder and JMeter from open source on performance testing.

There are various kinds of keywords which are handled in this technique. These are basically item or base level keywords, utility function and sequence or user keywords. In the keyword-driven testing also the keywords controlling the test execution are taken out of the scripts into the test data. This makes it possible to create new test cases in the test data without creating a script for every different test case allowing also the test engineers without coding skills to add new test cases. This removes the biggest limitation of the data-driven testing approach.

Table 2 shows an example of keyword-driven test data containing a simple test case for testing a login web application. The test cases consist of keywords Runapp, Username, Password and ok, and the arguments which are inputs and expected outputs for the test cases. As it can be seen, it is easy to add logically different test cases with-out implementing new keywords.

To be able to execute the tabular format test cases shown in table 3, there have to be mapping from the keywords to the code interacting with system under test (SUT). The scripts or code implementing the keywords are called handlers.

In Figure 2 can be seen the handlers for the keywords used in test data (table 2). In addition to the handlers, test execution needs a driver script which parses the test data and calls the keyword handlers according to the parsed data. If there is a need for creating high level and low level test cases, different level keywords are needed. Simple keywords like Username are not enough for high level test cases. There are simple and more flexible solutions.

Higher level keywords can be created inside the frame-work by combining the lower level keywords. The limitation of this approach is the need for coding skills whenever there is a need for new higher level keywords.

A more flexible solution proposed is to include a possibility to combine existing keywords in the keyword-driven test automation framework. This makes it possible to create higher level keywords by combining existing keywords inside the test data. These combined keywords as user keywords. [5]

TABLE II. KEYWORD-DRIVEN TEST DATA FILE

#	TYPE	KEYWORD	OPERATION	PARAMETER
1	Function	Runapp		http://in.yahoo.com/?p=us
2	Item	Username	SetValue	Username=rashmi
3	Item	Password	SetValue	Password=12345
4	Item	ok	Click	

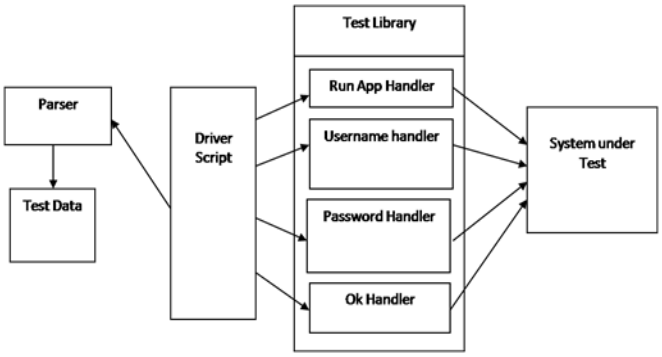


Figure 2. Handlers for keywords in Table 2

There are various advantages of using keyword driven testing techniques. These advantages are as follows:

1) Keywords that reflect the business can be chosen
2) Keyword re-use across multiple test cases
3) Not dependent on Tool / Language
4) The keyword list is robust to minor changes in the software.
5) Division of Labor Test Case

III. BASE REQUIREMENTS

There are various requirements that are known as "base requirements". These requirements must be fulfilled for success of keyword driven testing. These include:

1) The process of Test development and automation must be fully separated. It is very important to separate test development from test automation. Testers are not and should not be programmers. So, Testers must be adept at defining test cases independent of the underlying technology to implement them. Individuals who are skilled technically, the automation engineers, will implement the action words and then test them.
2) The test cases must have a clear and differentiated scope. It is important that test cases have a clearly differentiated scope and that they not deviate from that scope.
3) The tests must be written at the right level of abstraction such as the higher business level, lower user interface level, or both. It is also important that test tools provide this level of flexibility.

IV. KEYWORDS

The Test Language is based on a dictionary, which is comprised of words (keywords) and parameters. A Test Case is a sequence of steps that tests the behavior of a given functionality or feature in an application. Unlike traditional test approaches, Test Language uses pre-defined keywords to describe the steps and expected results. Keywords are the basic functional sub-procedures for the test cases of the application under test. A test case is comprised of at least one keyword. [4]

There are three kinds of keywords. These are item or base level keywords, utility function and sequence or user keywords.

Item is an action that performs a specific operation on a given GUI component. For example set value “rashmi" in “User name" control, set value “12345" in “Password control result" field.

When performing an operation on a GUI item, the following parameters should be specified: Name of GUI item, what operation to perform and the values. Table 3 shows Item operations.

Utility Function is a script that executes a certain functional operation that is hard or ineffective to implement as a Sequence. For example: Runapp, closeapp. Table 4 shows Utility Functions.

Sequence is a set of keywords that produces a business process, such as “Login”. Sequence keyword is made by combining various items and utility function. It is recommended to collect frequently used functional processes such as login, addition of new records to the system as a sequence instead of implementing them as items in test cases. Table 4 shows Sequence keywords.

Parameters are additional information required in order to generate the test conditions. In most cases, parameters should be defined for the created keywords. For example: failed authentication by passing username with illegal password, number for mathematical calculation, etc. Table 6 shows Keywords & their associated parameters

Examples for sequence parameters: When the user wants to create a new user, the following syntax is used: create_user (rashmi, 6/12/2000,rashmi.1306@yahoo.com)

Some of the keywords may contain dozens of parameters. In order to simplify the test creation, all parameters should contain default values. The tester should be able to change each one of the default parameters according to the context of the test. For example, if the tester would like to create new user that is older the 100 years, only the birth date will be changed and all other parameters will remain the same. Obviously, a specific change will not affect the default parameters being used for other tests.

TABLE III. ITEM OPERATION

#	TYPE	KEYWORD	OPERATION	PARAMETER
1	Item	Username	SetValue	Username=rashmi
2	Item	Password	SetValue	Password=12345

TABLE IV. UTILITY FUNCTION

#	TYPE	KEYWORD	PARAMETER
1	Function	Runapp	http://in.yahoo.com/?p=us
2	Function	Closeapp	http://in.yahoo.com/?p=us

V. OBJECT REPOSITORY

Object Repository is a centralized place for storing Properties of objects available in Application under Test (AUT). All software applications and websites are getting developed using many different components or small units like textbox control, input tag, web browser control etc. These components or small unit are known as Objects. [10]

Each object will be identified based on the object type. Each object will also have its own properties like name, title, caption, color, size. These properties help in the identification of these objects uniquely. There are also specified set of methods for each object. There are various properties that can be changed during run-time. These properties are known as Runtime Object (RO) properties. There are also some other properties that can't be changed. These properties are known as Test Object (TO) properties. [10]

There are some additional properties such as index, location which are known as ordinal identifiers. Actually these properties won't be available in the object of Application under the test. These are created in order to distinguish two objects which are having exactly same Test Object properties. For instance, some forms in the web pages will be have two submit buttons, one at top and another at bottom. These both can be identified separately on the basis of the location or index. As Test Object properties are also based on properties of object of Application under Test, there is no need for all the Test Object properties to be available in Runtime Object properties collection also. The object repository must support the editing of the properties of these Test Object and new properties can also be added to them. The value for the properties of the Test Objects in Object Repository need not be a constant. They can parameterize the values so that the Test Object property can be dynamically changed for each execution.

These properties are stored in the centralized place in object repository. This helps in the maintenance and updating of Test scripts can be easily done whenever there is a change in UI (User Interface) of the AUT. Assume that Login screen is used in around 20 Test scripts. If the Page name of login screen in changed, there is no need to make any change in all these 20 Test scripts. By just changing the property of Test Object in Object Repository is enough. A clear understanding of Object Repository is essential to carry out the operation of the Keyword driven testing successfully.

A framework for testing should be able to recognize any control or object in any webpage that needs to be testes. For recognizing the object, it should know the properties of those objects beforehand. During the execution of the test scripts, this identification is done. A framework has data tables for supporting the execution of multiple iterations of same step with different data.

There can be various methods to manipulate the test object properties. The Test Object properties of Test Objects can be accessed by implementing methods such as getTOproperty and getTOproperties.

TABLE V. SEQUENCE

#	TYPE	KEYWORD	PARAMETER
1	Sequence	Login	Username=rashmi Password=12345

TABLE VI. KEYWORDS & THEIR ASSOCIATED PARAMETERS

#	Type of keyword	Parameter
1	Item	Value
2	Function	Value
3	Sequence	Item

Even, Test Object property of Test Object can be changed using setTOproperty. It will be valid only till exiting the execution. After completing the execution it will resume the actual Test Object property stored in the Object Repository. During run-time we can get the property of the runtime object using getROproperty.

VI. KEYWORD DRIVEN MODULE

A framework used for performing keyword driven testing will consists of various interrelated modules. These modules are namely core module, scripting language module, support library module and many more. [1] Figure 3 shows the Keyword driven module.

Core module takes a major role in analysis the keyword information of the script, and controls the implementation of the scripts. It is composed of four parts that are the data parser, the script parser, the script actuator and the middle layer. The data parser is responsible for analysis on the keyword, the script parser responsible for analysis the logic keyword in the script, the script actuator is responsible for the implementation of the script, and the middle layer is responsible for calling the test. [1]

Data Access module is responsible for data storage, including add scripts, modify scripts, read scripts, enquiry scripts, delete scripts, and other functions. The script has three levels, when the high level and low-level script bearing the script, the layer will maintain this relationship. [1]

Interface module is to enhance the framework's ease of use. It realize a GUI interface, the graphical interface allows users to edit, drag and drop the modalities script; provides a user friendly guide to understand and use; provides view and editor which make it easily for users to view, modify the existing test scripts. [1]

Support module consists of two parts: one is the libraries that all of the tests can be shared, including the log library and the test supporting library. The log library is responsible for providing the functions of log records to testers; the test supporting library provides the functions that all of the tests can be shared. The second is the testing library for GUI; this part provides the controls libraries. [1]

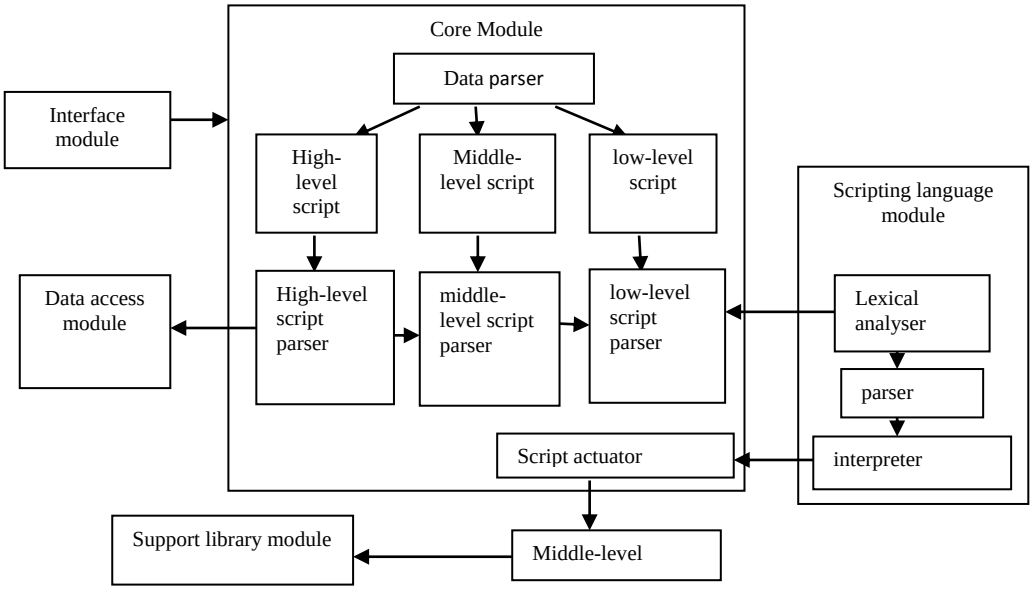


Figure 3. Keyword driven module [1]

Interpret scripting language module is consisted of three parts, such as lexical analyzer, parser, and interpreter. Lexical analyzer will be responsible for the output of the characters in a flow analysis as a word, parser will be responsible for the sequence of words with semantic analysis for the phrase and interpreter will be responsible for the semantic translation. [1]

VII. PPOPOSED APPROACH

First, a web application is taken which is needed to be tested. A web application can be a login page, online shopping web application, online reservation web application etc. To start the process of recording, enter the URL of the desired web application. The process of authentication in Railway Reservation web application, in which user passes user ID and a password, is taken as an example to describe the proposed approach.

As user navigates the Web application, the Keyword driven testing framework records the steps. In the User Name and Password boxes, the name and password are entered and the Sign-In Button is clicked. The travel planning web page opens (If both correct username and password are entered). These operations form the basis of the test. The tool records all the operations performed in the Web browser until the recording is stopped.

When the recording is stopped, a test script file is generated. This generated test script file containing all user actions, is saved. The user actions will comprise of items clicked, items selected, value typed etc, during the recording procedure. The tool will also generate steps in the table format, representing each operation performed in the form of Keyword, Value and Operation. For example, User Name text field, Password text field and Sign-In button are the keywords. The Value represents values entered by the user in the items. The Operation includes the click, select, drag or drop, etc. The test Script will be useful in the play back of a test and reusability of the test script i.e. parameterization and multiple actions.

When the test is play backed, the tool runs the saved test script file. The recorded web application opens in the web browser and all steps are performed automatically, as it was originally recorded in the test. For example, the recorded test script for authentication process can be played backed. Parameterized tests and multiple action tests are also played back using the play back module.

When the test run is completed, it displays the results of the run (whether a test is passed or failed) in the test result page. The Test Results window opens, which contains the result summary of the test execution. The Test Results window displays the key elements of the test run for test analysis purpose. The key elements are composed of two parts. First element shows the steps (in the tree structure format) that were performed while the test was running. The second element is the test result details. The test result contains iteration and status summary. The iteration summary indicates which iterations passed and which failed. The status summary indicates the number of test or reports that passed, failed, and raised warnings during the test.

Object Repository is a centralized place for storing the properties of objects available in AUT (Application under Test). The keywords can be added in the object repository, either manually or at the time of recording. All software applications and websites are developed using many different components or small units (for example *textbox control in VB, input tag in HTML, web browser control in .net*) which are known as Objects. Each object is identified on the basis of the object type. Each object has properties (for example *name, title, caption, color and size*) and specific set of method, which help in identification of an object. The object repository will support the modification of Test Object's properties, as well as, new properties can also be added. The values of the properties stored in Object Repository need not be a constant. The values can be parameterized by making them variable.

The Test scripts can be added, deleted and modified. The test script modification helps in the parameterization of the test

as well as during the division of a test into multiple actions. While testing a web application, there may be needed to check how the web application performs the same operations with multiple sets of data. For example, how a Web application responds to twenty separate sets of data is to be checked. There are two ways to do this. Twenty separate tests are recorded, each with its own set of data. Here, no reusability of Test Script. Alternatively, A test is recorded, and the values of the objects in this test are made variable for parameterization. This saves the nineteen runs of the recording process. This single test with the help of parameterization can be played twenty times using a different set of data each time. Each test run is called iteration. All iterations are numbered. Later is the better approach and involves the reusability of Test Script. In the above example, the authentication page is signed in with 'rashmi.1306' as user ID and 'dracoXXXXXX' as password. The 'rashmi.1306' is a constant value, which means that 'rashmi.1306' is the user ID each time the test is run. Through the data table parameter, the user ID can be changed into a variable, so that a different user id can be used for each test run. In the parameterized Keyword user ID, two different user ids can be added for example 'ras_gupta' and 'prati_gupta'. The tool can be test the application with this different data without the need of recording with these data.

Actions divide the test into various logical sections. When a new test is created, it is represented as a single action. By dividing the tests into multiple actions, more modular and efficient tests can be designed. This is another feature of reusability of keywords that makes Keyword driven Testing more efficient and modular than the Data Driven Testing. To explain this we take the whole example of the above railway reservation web application that books a flight. This can be divided into several distinct processes or actions which are as follows:

- The Online railway reservation web application is logged in.
- The trains are booked.
- Another action for logged out from the web application.

The above test can also be parameterized for ten different train booking. This parameterized test now can be run ten times using ten different sets of data. With the help of Multiple Action, the test can also be organized so that only the second procedure runs ten times, simulating a single user logging in, booking ten trains, and logging out. This can be done by dividing the test into different actions. This saves the nine runs of Logged-in and Logged-out process.

The parameterized tests and tests divided into multiple actions are also play backed simply and their test results are analyzed. Figure 4 shows the above proposed approach in a flow chart.

VIII. CONCLUSION

In this paper, the different types of keywords, base requirement, methodology, object repository and various keyword driven modules are investigated. These all are required to carry out a successful and efficient operation of Keyword driven testing. It is important to understand that keywords are not magic, but they can serve well. It is essential to do test design in a right and efficient way. The process of the test automation should be done but it should not dominate the process. It should flow from the overall strategy, methodology, and architecture. Moreover, the existing tools available for this approach make use of the HTML, Xml, spreadsheets to maintain test cases in object repository which are not very scalable.

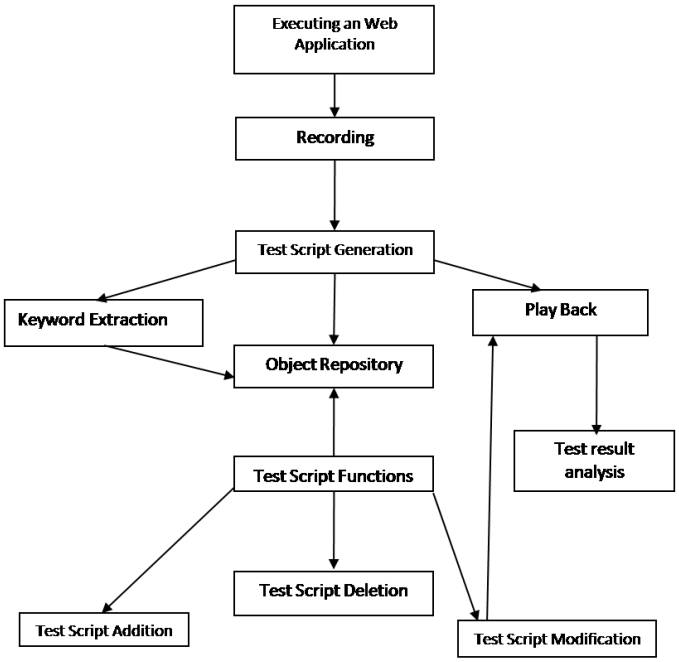


Figure 4. Proposed Approach

REFERENCES

[1] Jie Hui, Lan Yuqing, Luo Pei, Gao Jing, Guo Shuhang, "LKDT: A Keyword-Driven Based Distributed Test Framework", International Conference on Computer Science and Software Engineering, 2008, pp. 719-722

[2] Pekka Laukkanen, "Data-Driven and Keyword-Driven Test Automation Frameworks", Helsinki University of Technology, Department of Computer Science and Engineering Software Business and Engineering Institute. 2007, pp. 1-102

[3] Juha Rantanen,"Acceptance Test-Driven Development with Keyword-Driven Test Automation Framework in an Agile Software Project" Helsinki University of Technology, Department of Computer Science and Engineering, Software Business and Engineering Institute. 2007, pp. 1-102

[4] Ayal Zylberman and Aviram Shotten, "Test Language: Introduction to Keyword Driven Testing". 2010, pp 1-7

[5] Tommi Takala, Mika Maunumaa, and Mika Katara. "An Adapter Framework for Keyword-Driven Testing", Department of Software Systems, Tampere University of Technology, Finland. Ninth International Conference on Quality Software. 2009, pp. 201-210

[6] http://en.wikipedia.org/wiki/Keyword-driven_testing

[7] Bharath Anand R., Harish Krishnankutty, kaushik Ramakrishnan, Venkatesh V.C.," Business Rules- Based Test Automation- A novel Approach for accelerated testing". 2007, pp. 1-12

[8] Liu Xing, Li Yan, Cai Mian, Guo Ying, "The Testing and Evaluation System for the Secure Operating System Based on the Mechanism of

keyword-driven". Ninth International Conference on Information Assurance and security. 2009, pp. 471-474

[9] http://en.wikipedia.org/wiki/Test_automation

[10] http://www.automatedqa.com/products/testcomplete/ manager-overview/

[11] Bennett, "J.P. Introduction to Compiling Techniques – A First Course Using ANSI C", Lex and Yacc [M]. McGraw Hill Book Co, 1990.

[12] Nancy S. Eickelmann, "An evaluation of software test environment architectures". International Conference on Software Engineering, 1996, pp. 353-364.

[13] Terence Parr," The Definitive ANTLR Reference: Building Domain-Specific Languages", Pragmatic Bookshelf, 2007, pp 14-85.

[14] Sheng Liang. Java(TM) Native Interface, "Programmer's Guide and Specification", Prentice Hall PTR, 1999, pp 1-35.

[15] Mercury QuickTest Professional Tutorial, Version 8.0

[16] Kaner, "Pitfalls and strategies in automated testing", IEEE Computer, 30(4): April 1997, pp 114–116,

[17] Kaner, J. Bach, and B. Pettichord, "Lessons Learned in Software Testing: A Content-Driven Approach", John Wiley & Sons, Inc., 2001.

[18] Kelly, " Choosing a test automation framework, July 2003",URLhttp://www106.ibm.com/developerworks/rational/library/591.html. April 30, 2005.

[19] Kit, "Integrated, effective test design and automation". Software Development, February 1999, pp 27–41.

AUTHORS PROFILE

Rashmi, done B.Tech (Computer Science) in 2010 with 79% from M.D.U. (Rohtak), Currently pursuing M.Tech (Computer Science) from Centre for Development of Advanced Computing, Noida, I.P. University and doing a project on "A Keyword driven framework for testing web applications".

Mrs. NEHA BAJPAI received M.Tech in Information Technology from the Vinayaka Mission University of Tamilnadu in the year 2005. She has ten years of teaching and one year of IT implementation experience. Presently, she is working as a Senior Faculty in School of IT at Centre for Development of Advanced Computing (CDAC), Noida. Her present interests are in the subjects related to Object Oriented Technologies, Oriented Analysis & Design, UML, Software Testing and Object Oriented Database Management System. She has over 15 research papers in various international and national Journals, Conferences & Seminars. She also served, coordinated and taught various International Training Programs under Indo-Vietnam bi-lateral Cooperation and ITEC/SCAAP Scheme of MEA.

Effect of Error Packetization on the Quality of Streaming Video in Wireless Broadband Networks

Aderemi A. Atayero
Department of Electrical and Information Engineering
Covenant University
Ota, Nigeria

Oleg I. Sheluhin and Yury A. Ivanov
Department of Information Security
Moscow Tech. Univ. of Communication and Informatics
Moscow, Russia

***Abstract*—A Markov model describing the duration of error intervals and error-free reception for streaming video transmission was developed based on the experimental data obtained as a result of streaming video from a mobile source on IEEE 802.16 standard network. The analysis of experimental results shows that the average quality of video sequences when simulating Markov model of packetization of errors are similar to those obtained when simulating single packet errors with PER index in the range of 3×10^{-3} to 1×10^{-2}. An algorithm for creating software for simulating packetization of errors was developed. In this paper we describe the algorithm, software developed based on this algorithm as well as the Markov model created for the modeling.**

Keywords-Video streaming; Markov model; IEEE 802.16; Bit Error Rate; Burst Error Length; Packet Error Rate; Codec.

I. INTRODUCTION

The need to create realistic simulation and mathematical models of behavior of losses in the communication channels based on the apparatus of Markov chains for wireless access systems is a scientific problem of important consequence. Markov processes with the necessary number of states sufficiently describe the mechanism of transmission of information [1], the knowledge of which is necessary to analyze network problems during packet video transmission. The parameters of the model make it possible to determine the quality of transmitted video as well as the statistical parameters of the network.

A model decscribing the length of error intervals and error-free reception for streaming video transmission was developed based on the experimental data obtained as a result of streaming video from a moving source on WiMAX network [2]. Based on the graph of packet loss distribution, an array was formed in which the lost packet corresponds to a logic zero (0) and received packet corresponds to a logic unit (1). The original array was split into two, one of which contains information about the lost packets and the other contains information about the received packets. The formation of arrays was carried out in accordance with the procedure shown in Figure 1.

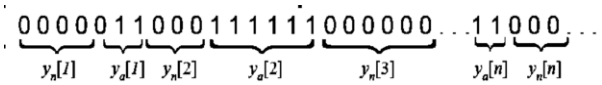


Figure 1. Formation of Arrays

II. MARKOV MODEL DESCRIBING THE EXPERIMENTAL DATA

In accordance with the method presented in [3], the available raw data file was divided into two parts, each of which separately contains the duration of ON periods and OFF periods. Variables $y_a[n]$ fall under the ON periods, while variables $y_n[n]$ fall under the OFF peroids. An approximation of the distribution function (DF) of real processes is obtained. Equation (1) is used for approximating the distribution function of OFF state.

$$F^*(k) = A_i \sum_{i=1}^{3} e^{-\alpha_i k} \quad (1)$$

By using the method of least squares we find the unknown coefficients of the approximation for the expression (1) as presented in Table 1.

TABLE I. APPROXIMATION COEFFICIENT VALUES (A_i ; α_i)

A_1	α_1	A_2	α_2	A_3	α_3
0.612086	0.072672	0.631933	0.540023	0.073586	0.040006

Substituting the coefficient values obtained and given in Table 1 into equation (1), we obtain the approximation of the original distribution of the length of OFF periods as equation (2):

$$F^*(k) = 0.612086 \times e^{-0.072672k} + 0.631933 \times e^{-0.540023k} + 0.073586 \times e^{-0.040006k} \quad (2)$$

Equation (3) is used for approximating the distribution function of ON state.

$$F^*(k) = B_i \sum_{i=1}^{6} e^{-\beta_i k} \quad (3)$$

The unknown coefficients of the approximation for the expression (3) are found using the method of least squares and presented in Table 2.

TABLE II. APPROXIMATION COEFFICIENTS (B_i ; β_i)

B_1	β_1	B_2	β_2	B_3	β_3
0.065836	0.000643	0.107716	0.000708	0.33109	0.007203
B_4	β_4	B_5	β_5	B_6	β_6
0.057449	0.0000618	0.007203	0.291568	0.224767	0.0094038

Substituting the coefficient values obtained and given in Table 2 into equation (3), we obtain the approximation of the original distribution of the length of ON periods as equation (4):

$$F^*(k) = 0.065836 \times e^{-0.000643k} + 0.107716 \times e^{-0.000708k} + 0.33109 \times e^{-0.007203k} + 0.057449 \times e^{-0.0000618k} + 0.007203 \times e^{-0.291568k} + 0.224767 \times e^{-0.0094038k} \quad (4)$$

The approximation of DF of ON is shown in Figure 2.

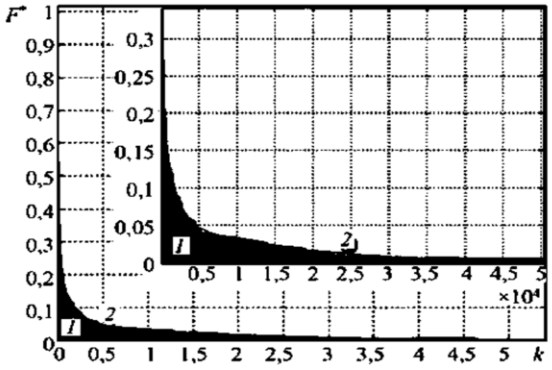


Figure 3. DF of ON approximation (2), DF of ON experiment (1). (embedded graph - reduced scale of DF of ON)

After the normalization of obtained approximating expressions (2) and (4), additional distributions of duration of ON-and OFF-processes, the matrix of transition probabilities are created, which is of the form presented in Figure 3:

$$\Gamma = \begin{array}{c|cccccc} & A_1 & A_2 & A_3 & B_1 & B_2 & B_3 \\ A_1 & e^{-\alpha_1} & 0 & 0 & (1-e^{-\alpha_1})B_1 & (1-e^{-\alpha_1})B_2 & (1-e^{-\alpha_1})B_3 \\ A_2 & 0 & e^{-\alpha_2} & 0 & (1-e^{-\alpha_2})B_1 & (1-e^{-\alpha_2})B_2 & (1-e^{-\alpha_2})B_3 \\ A_3 & 0 & 0 & e^{-\alpha_3} & (1-e^{-\alpha_3})B_1 & (1-e^{-\alpha_3})B_2 & (1-e^{-\alpha_3})B_3 \\ B_1 & (1-e^{-\beta_1})A_1 & (1-e^{-\beta_1})A_2 & (1-e^{-\beta_1})A_3 & e^{-\beta_1} & 0 & 0 \\ B_2 & (1-e^{-\beta_2})A_1 & (1-e^{-\beta_2})A_2 & (1-e^{-\beta_2})A_3 & 0 & e^{-\beta_2} & 0 \\ B_3 & (1-e^{-\beta_3})A_1 & (1-e^{-\beta_3})A_2 & (1-e^{-\beta_3})A_3 & 0 & 0 & e^{-\beta_3} \end{array}$$

Figure 4. The matrix of transition probabilities

Substituting the values of the coefficients found in Tables 1 and 2 into the matrix of transition probabilities, we obtain the matrix of values in Figure 4.

$$\Gamma = \begin{vmatrix} 0.999 & 0 & 0 & 2.4*10^{-5} & 8.69*10^{-4} & 1.1*10^{-4} \\ 0 & 0.9944 & 0 & 1.344*10^{-4} & 0.0049 & 6.16*10^{-4} \\ 0 & 0 & 0.965 & 8.4*10^{-4} & 0.0304 & 0.0039 \\ 1.8*10^{-5} & 3.6*10^{-4} & 3.6*10^{-4} & 0.9991 & 0 & 0 \\ 4.2*10^{-5} & 8.4*10^{-4} & 8.4*10^{-4} & 0 & 0.9979 & 0 \\ 2.58*10^{-4} & 0.0052 & 0.0052 & 0 & 0 & 0.9871 \end{vmatrix}$$

Figure 2. The matrix of values

III. SOFTWARE FOR ERROR PACKETIZATION SIMULATION

Simulation of the transmission of streaming video traffic over a WiMAX network can be done given the probability transition matrix and vector of initial probabilities [4]. The choice of the initial state of the system was carried out using the condition that all states are equiprobable (i.e. $p = 1/N$, where N-number of states the system can be in after DF approximation). Description of the block diagram of the simulation algorithm is as given below, while the Markov model is shown in Figure 5.

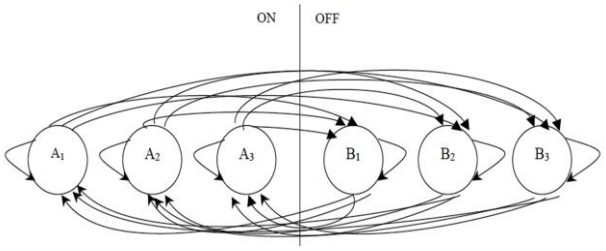


Figure 5. Markov model of error packetization algorithm

IV. DESCRIPTION OF ERROR PACKETIZATION ALGORITHM

STEP 1. Start program (Description of the variables, functions, procedures and modules used)

STEP 2. Enter two-dimensional array matrix of transition probabilities. In the developed software, this matrix was given as an array of constants in the declarations section and named *markov*.

STEP 3. Set state from which to begin modeling. Since a 9-state model was chosen, the state variable can take integer values on the interval (1 – 9). Also, at this stage of the algorithm the accumulated variables *summa_on* and *summa_off*, which reflect the duration of the ON periods and OFF periods, are reset to zero respectively.

STEP 4. Begin cycle with parameter *i*. The number of iterations equals the number of transitions in the simulated system.

STEP 5. Instantiate the built-in generator of pseudorandom uniformly distributed sequence, generating a random value in the interval (0, 1). Assign the generated value to *rnd*. At the moment of generating the variable *rnd*, the system moves to the next state. The

exact state into which it falls will be determined by the subsequent actions of the algorithm. The variable *summa* is reset to zero.

STEP 6. Start the cycle with parameter *k*. The number of iterations in the cycle equals the number of states of the system being modeled. For this case, the number of iterations is eight (8). This loop is used to determine the state into which of the system has moved at the particular time of consideration.

STEP 7. Check – does the value of *rnd* fall in the k^{th} state of the Markov chain. At the same time the following variables are involved: summa - accumulates the probability of all states up to the k^{th}; markov [state, k] – a two-dimensional array, which contains the transition matrix. If rnd falls within a range of probabilities corresponding to the k^{th} state, then goto step 8, otherwise goto step 9.

STEP 8. Check – in which state is the process currently? If in the active state, then goto step 10. If in passive state, then goto step 11.

STEP 9. The summa variable is increased by the value of the probability of being in state k. Then proceed to the next iteration of step 6.

STEP 10. Check – was the last state of the matrix passive? If yes, goto step 12. Otherwise, goto step 16.

STEP 11. Check – was the last state of the matrix active? If yes, goto step 13. Otherwise, goto step 17

STEP 12. Arrival at this step implies the end of OFF period. Therefore save or print to file *summa_off*.

STEP 13. Arrival at this step implies the end of ON period. Therefore save or print to file *summa_on*.

STEP 14. Since the OFF period as ended, reset the variable *summa_off* to zero in preparation for the record of fresh OFF-period information, when the process will be in the passive state.

STEP 15. Since the ON period as ended, reset the variable *summa_on* to zero in preparation for the record of fresh ON-period information, when the process will be in the active state.

STEP 16. Arrival at this step implies either the continuation of the previous ON period, or the start of a new ON period. So increment the variable *summa_on* and assign the value of cycle *k* to the *state* variable.

STEP 17. Arrival at this step implies either the continuation of the previous OFF period, or the start of a new OFF period. So increment the variable *summa_off* and assign the value of cycle *k* to the *state* variable.

STEP 18. At this step of the algorithm, the system just transited to the next state, so turn to the next iteration of the parameter *i*.

STEP 19. End program.

As a result, the amount of packets falling either in the received state or the lost state in a row is accumulated ($summa\_ON = summa\_ON + 1$).

V. DISCUSSION

Distribution function of ON- and OFF-processes for both the simulated and experimental sequences was obtained using the described Markov model shown in Figure 6.

Experiments show that increasing the number of states of the Markov model describing the packetization of errors allows for obtaining a satisfactory correspondence between the results of the experimental data and the data obtained by simulation.

A. Markov Model of Packetization of Errors

Two independent datasets, each containing 300,000 values were generated with the aid of the developed Markov model [5]. This amount of data allows for a qualitative comparison of RTP packets from the experiment conducted on the transmission of a 30-minute streaming video on a real WiMAX network [2] with the results of the experiment conducted using the HSC.

In the model, each value in the array is represented by the numbers 0 or 1, where 0 means error-free value, and 1 – erroneous value. Figure 7 shows the distribution of data set values, where the white areas correspond to error-free values (0), and black - erroneous values (1).

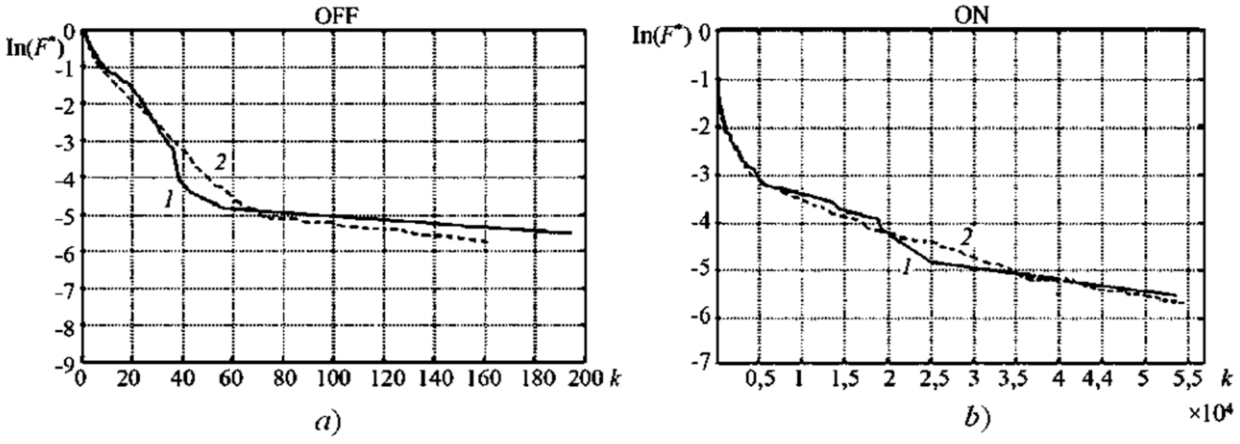


Figure 6. DF of simulated samples of the length of OFF-(a) and ON (b) - periods: curve 1 - experiment, curve 2 - simulation

Figure 7. Distribution of error-free and erroneous values for arrays №1 and №2.

The first array contains 2,743 (0.91%), and the second has 2,430 (0.81%) errorneous values. The distribution of the number of errors in the same error group is presented in the form of histograms in Figure 8.

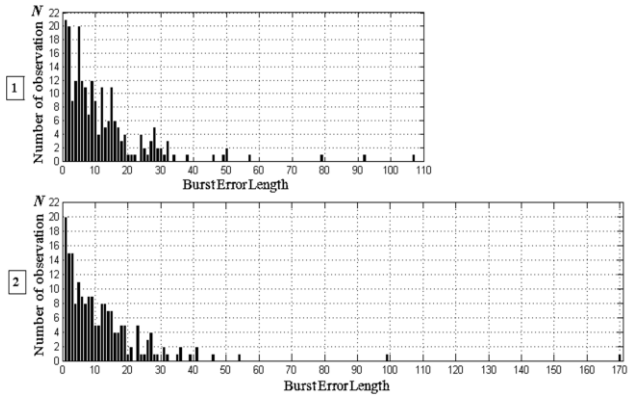


Figure 9. Distribution of errors in a group of bugs array №1 and №2.

It is shown that the distribution of errors cannot be approximated by an exponential function, a fact that validates the use of the Markov model. Furthermore, in order to study the influence of the Markov model of packetization of errors on the quality of video streaming, simulation of the transmission of a 30-minute video in the structure of the HSC was conducted. The simulation entailed the transmission and reception of traces over an "ideal" channel with unlimited bandwidth and no delay in the NS-2 environment [6]. Subsequently each packet of the receive trace was matched with a corresponding value from the dataset array (packet id = serial value of the data set array). All packets corresponding to 1 (indicating error) were deleted. This allowed for simulating sequence of errors that occur in the network and to effect corrective decoding of the video stream.

Two experiments were carried out with arrays №1 and №2 respectively. Figure 9 shows a block diagram of the experiments. The results of experimental quality indicators obtained are shown in Figures 10 – 12.

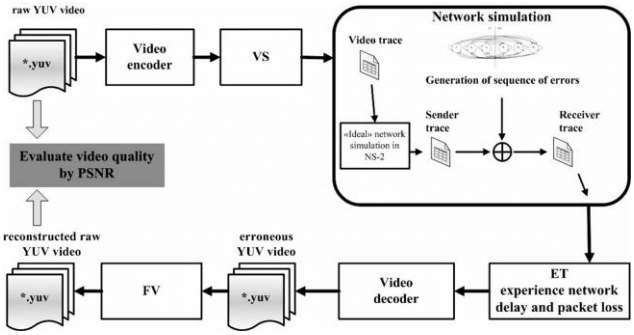


Figure 8. Block diagram of the experiments №1 and №2.

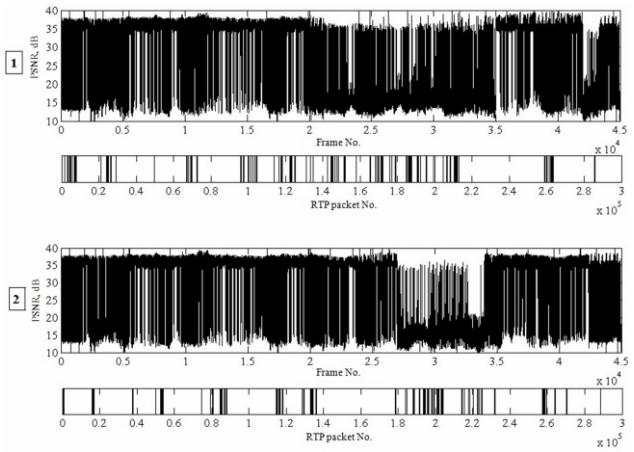


Figure 10. The change the PSNR indicator from experiments №1 and

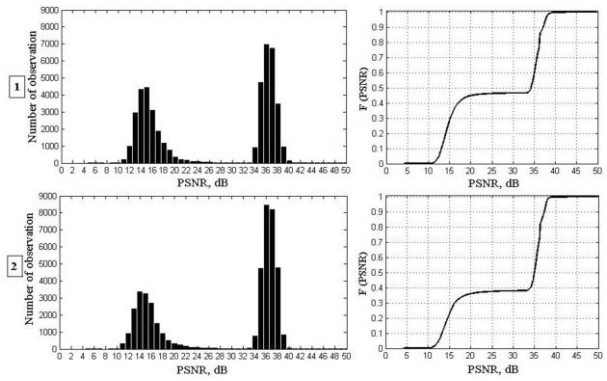


Figure 11. Histogram and distribution function of the PSNR indicator in experiments №1 and №2

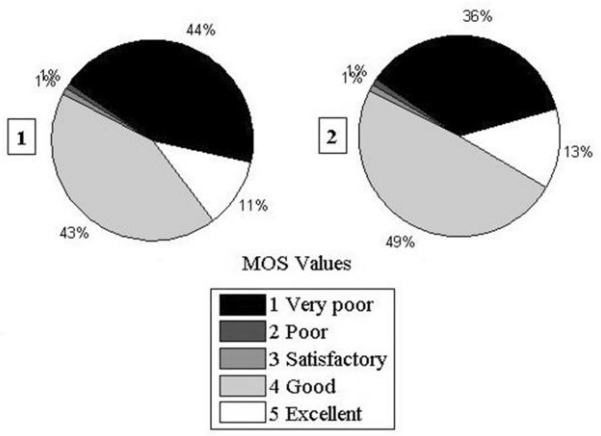


Figure 12. MOS Quality value for video broadcast in №1 and №2

VI. CONCLUSION

Analysis of the quality of received video sequence when simulating Markov model of error packetization shows that the average quality of video sequences is slightly worse than during transmission over a real network. For example, in an experiment on streaming video over a real WiMAX network, the average quality of 31 dB was obtained, and for the simulations 26 dB and 28 dB respectively. The subjective MOS quality indicator also shows a difference in values: a real WiMAX network returned a mean value of 3.59 (corresponding to satisfactory), while the experiments returned values of 2.72 (corresponding to poor) and 3.01 (corresponding to satisfactory), respectively. This suggests that the Markov model of packetization of error obtained from a real network for streaming video can be used in the simulaton of transmission of video across networks in the hardware-software complex developed by the authors in a previous work [7].

The average quality of video sequences obtained from simulations of the Markov model are similar to those obtained when simulating single packet errors with PER index in the range of 3×10^{-3} to 1×10^{-2}. While the length of error group depending on the PER of the specified range can attain values of BEL ≤ 10.

REFERENCES

[1] H. Wang and N. Moayeri, "Finite state Markov channel - a useful model for radio communication channels," IEEE Trans. on Vehicular technology, vol. 44, № 2, pp. 163-171, February 1995.

[2] Atayero A.A., Sheluhin O.I., Ivanov Y.A. and Iruemi J.O., "Effect of wideband wireless access systems interference robustness on the quality of video streaming," Lecture Notes in Engineering and Computer Science: Proceedings of The World Congress on Engineering and Computer Science 2011, WCECS 2011, 19-21 October, 2011, San Francisco, USA, pp. 848-854.

[3] ITU P.800: Methods for subjective determination of transmission quality, available at: http://www.itu.int/rec/T-REC-P.800-199608-I/en.

[4] Deb S., Jaiswal S. and Nagaraj K., "Real-time video multicast in WiMAX networks," Proc. of IEEE INFOCOM, April 2008.

[5] Hohlfeld O., "Markovian packet loss generators and video QoE, T Systems, February 2008.

[6] NS-2 Documentation, Available at: http://www.isi.edu/nsnam/ns/ns-documentation.html, last accessed 11/11/2011.

[7] A.A. Atayero, O.I. Sheluhin, Y.A. Ivanov, A.S. Alatishe "Estimation of the Visual Quality of Video Streaming Under Desynchronization Conditions," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 2, № 12, pp. 1-11, December 2011.

AUTHORS PROFILE

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

He is a member of a number of professional associations including: the Institute of Electrical and Electronic Engineers, IEEE, the International Association of Engineers, IAENG, and a professional member of the International Who's Who Historical Society (IWWHS) among others. He is a registered engineer with the Council for the Regulation of Engineering in Nigeria, COREN. A two-time Head, Department of electrical and Information Engineering, Covenant University, Nigeria, he was the coordinator of the School of Engineering of the same University.

Dr. Atayero is widely published in International peer-reviewed journals, proceedings, and edited books. He is on the editorial board of a number of highly reputed International journals. Atayero is a recipient of several awards including the '2009/10 Ford Foundation Teaching Innovation Award'. His current research interests are in Radio and Telecommunication Systems and Devices; Signal Processing and Converged Multi-service Networks.

Oleg I. Sheluhin was born in Moscow, Russia in 1952. He obtained a M.Sc. Degree in Radio Engineering1974 from the Moscow Institute of Transport Engineers (MITE). He later enrolled at Lomonosov State University (Moscow) and graduated in 1979 with a Second M.Sc. in Mathematics. He received a PhD at MITE in 1979 in Radio Engineering and earned a D.Sc. Degree in Telecommunication Systems and Devices from Kharkov Aviation Institute in 1990. The title of his PhD thesis was 'Investigation of interfering factors influence on the structure and activity of noise short-range radar'.

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

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

Yury A. Ivanov was born in Moscow, Russia in 1985. He obtained a M.Sc. degree in Systems, network and devices in telecommunications from Chuvash State University in 2007. He obtained a Ph.D in Telecommunication Networks and Systems in 2011 from Moscow State University of Communication and Informatics. Dr. Ivanov has published over 35 scientific papers and his current research interests include Radio and Telecommunications Systems and Devices: transmission of multimedia data across telecommunication networks, assessment of the quality of video sequences. Dr. Ivanov is a member of a number of professional associations including: the Institute of Electrical and Electronic Engineers, IEEE, the International Association of Engineers, IAENG.

An Overview of Video Allocation Algorithms for Flash-based SSD Storage Systems

Jaafer Al-Sabateen [1], Saleh Ali Alomari [2] and Putra Sumari [3]
[1,2,3] Multimedia Computing Research Group, School of Computer Science
University Sains Malaysia
11800 Pulau Pinang, Malaysia

***Abstract*—Despite the fact that Solid State Disk (SSD) data storage media had offered a revolutionary property storages community, but the unavailability of a comprehensive allocation strategy in SSDs storage media, leads to consuming the available space, random writing processes, time-consuming reading processes, and system resources consumption. In order to overcome these challenges, an efficient allocation algorithm is a desirable option. In this paper, we had executed an intensive investigation on the SSD-based allocation algorithms that had been proposed by the knowledge community. An explanatory comparison had been made between these algorithms. We reviewed these algorithms in order to building advanced knowledge armature that would help in inventing new allocation algorithms for this type of storage media.**

Keywords-SSDs; Allocation Algorithms; Data Management Systems; Garbage Collection; Storage Media.

I. INTRODUCTION

The Solid State Disk (SSD) is a high performance data storage device that hasn't any moving parts, and achieved a superior performance comparing with traditional storage media [3, 6, 7, 16]. It's based on flash memory technology [7, 20]. Nowadays, flash-based Solid State Drive (SSD) has been widely used as a storage device for many systems such as laptops computers, enterprise servers and other digital devices [3, 5, 11]. The widely used were caused by many features that are available in this type of storage devices. For example, these systems have low-power consumption features, non-volatility properties, high random access performance, and high mobility platforms [8, 11, 19, 22]. For several years, there has been a significant growth in the flash-based Solid State Drive (SSD) market, due to the previous mentioned features [15, 17]. Recently, due to the sequential price reduction of flash-based storage systems, the Solid State Drive (SSD) is emerging as a killer application NAND flash in general purpose computing areas such as, desktop personal computers [11, 17], and enterprise servers [1, 11]. Technically, there are two basic types of SSD storage systems, Multi-Level Cell (MLC), and Single Level Cell (SLC) [11, 17, 21]. The main differences between these two types are concerning with the number of writing cycles [8, 10], and the storage capacity level, which mainly affected the SSD media life span [10, 14, 21, 22]. Physically, there are three main components that SSD data storage system consists of, the flash package, SSD controller and host interface logic [7, 12, 14]. These components are simplified and illustrated in the following figure.

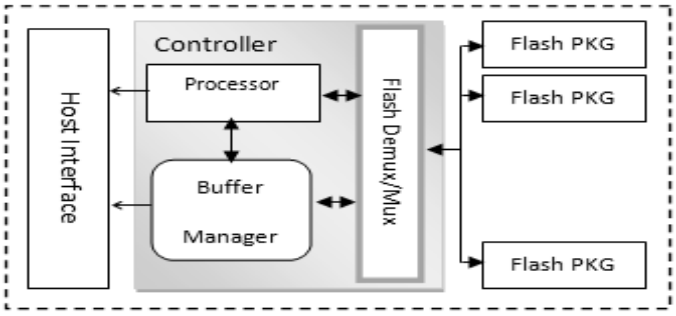


Figure 1. Basic Architecture of Solid State Disk (SSD) Storage Media

The controller as shown in figure 1, consist of three main parts, processor, buffer manager, and flash demux/mux integrated circuit [14]. The main function of the processor is to manage the flow of data and mappings from the logical block address to physical locations [5, 11, 14]. Respectively, the buffer manager will speed up the processing time required for performing several storage system functions, such as reading or writing operations [9, 10], while flash demux/mux is to managing instructions and data transport processes along the serial connections to the flash packages [13,15]. The host interface represents the point that connects internal environment for SSD storage system with the external physical environment. Regarding to package component, there are three basic terms should be clarified, Cell, Page, and Block components. The cell is the basic smallest unit of the block. The page consists of fixed number of cells, and a several sets of pages could be grouped into one block. An overall simplified architecture is shown in figure 2.

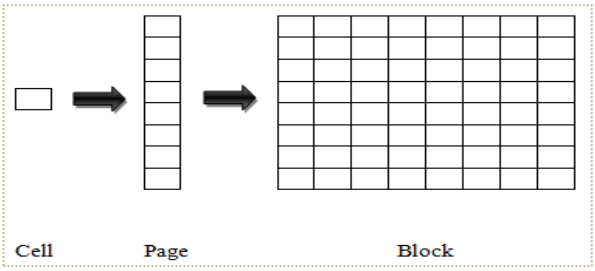


Figure 2. Block Basic Architecture

In this investigation paper, we will present a group of allocation algorithms that were proposed to handle some of challenges that emerged from three main Management issues in SSD storage systems, which are garbage collection process,

limited erasure cycles, and out-place updating method [2, 4, 8 and 16].

The remainder of this overview study is organized into five main sections. Section (2) is the related background of flash-based SSD storage disk, which presents the storage system and explains the basic components that are forming it, clarifying some operational characteristics that executed in it such as out place updating scheme, garbage cleaning process, and limited erasure cycles is included there also. Section (3) shows the motivation factors that motivate the researchers to design and implement a set of flash-based allocation algorithms and the core problem that this investigation paper focuses on. The section (4) provides a literature review for allocation algorithms proposed for flash-based SSD system and suggested in the knowledge community. Regarding to section (5), it presents a comprehensive comparison between these algorithms and shows the main Discriminated Degree between it, and finally, section (6) clarify the conclusions for this investigation paper and general recommendations.

II. Related Background

The flash-based Solid State Disk (SSD) consists of a big number of blocks [17]. The read and write of data is done on cell basis while the erase is carried on block unit [1, 3, 19, 20, 22]. There are three basic operational characteristics, out place updating scheme, garbage cleaning process, and limited erasure cycles [3, 7, 8, 11, 16, 20, 21]. In flash-based Solid State Disk storage media, updating the existing data by overwriting the same physical location is strictly prohibited [11, 20, and 21]. The original data must be erased prior to the updated data can be stored on same location. To avoid from initiating the erase operation each time the data is updated, out place updating scheme has been recognized in flash-based storage media [3, 7, 8, 11, 16, 2]. In this scheme, the update data is written in new location while the old version is marked as invalid [1, 3, 7, 8, 11, 16]. After a long series of updating transaction occurs in the media, big volume of available space in flash-based storage media is consumed [6, 17, 22]. The garbage cleaning process is invoked when the ratio of invalid (garbage) reaches a certain level of threshold [2, 17]. Before the process can be initiated, valid data resided in the block must be copied into available free spaces of other blocks [1, 5, 21]. In this situation, additional processing time is consumed by garbage cleaning (collection) process [4, 5, 13, 19]. Moreover, system resources are consumed in form of system performance weakness. The erase operation is necessary in flash-based memory systems, in order to ensure the continuity of data storing process. Limited erasure cycles refer to number of erase operation allowed to each block [6, 11]. Each block has its own limitation in erasure access lifespan [7, 9, 19, 21, 22]. Excessively accessing will cause the block become unreliable and spoiled [9, 12, 15]. These factors affected on the performance of this media [3, 11, 22]. In The midst of this processes in the storage media, the need for allocation algorithm had become a critical option.

III. Allocation Algorithms Motivations

The flash-based Solid State Disk (SSD) storage media have three hardware characteristics, garbage collection mode; out-place updating scheme and limited erasure cycles, which affected its general performance [3, 7, 8, 11, 16, 20, 21]. The space waste, time consumption, and Random writing operations, had been appeared as a group of performance challenges that emerged from these characteristics. An allocation algorithm that mitigates its ruggedness should be invented and presented.

IV. SSD-Based Allocation Algorithms

To eliminate the ruggedness of the out place updating scheme, garbage cleaning process, and limited erasure cycles. Many of flash-based allocation algorithms had been proposed in the literature. Table 1 show these algorithms which had been proposed by different authors, this contribution came as a group of relief options that should mitigate the ruggedness degree of the SSD operational characteristics.

TABLE I. Research Contribution For SSD-Based Allocation Algorithms

Algorithms	Publication	Author(s)
FCFS	2005	Li-Fu Chou, Pangfeng Liu[3]
FRFS		
On-line FRFS		
Probability-based	2009	Putra Sumari, Amir Rizaan Rahman [22]
Best-M	2007	Pangfeng Liu, Chung-Hao Chuang, and Jan-Jan Wu [19]

However, next sections reviews and discusses these proposed allocations algorithms in more details.

A. First Come First Serve (FCFS)

First Come First Serve (FCFS) [3] is a flash-based memory systems allocation algorithm. FCFS suggests the arrival time property for each incoming data page to be as base of allocation decision. FCFS allocates the modified data pages without any computation complexity. It uses a blocks list data structure which consider as memory blocks store. For the incoming modified data pages, FCFS algorithm selects the appropriate block from this list. In general allocation procedure of FCFS algorithm, the algorithm places the first modified data page into the first position of the first block of the selected block that had selected from block list data structure. The second page is placed into the second position of the first blocks of the selected block from block list data structure, and so on. FCFS allocation algorithm, places the data pages into blocks in sequentially style. The following Pseudo code explains how allocation decision is made during this algorithm:

1. *for each page in the page access pattern*
2. *Place the page into the first available cell from the block list.*
3. *If the page appeared before then:*
4. *mark the cell it previously resided INVALID*
5. *If the block the page previously resided now becomes INACTIVE then:*
6. *erase it and move it to the end of the block list*

Algorithm 1. Pseudo Code for First Come First Serve (FCFS) Allocation Algorithm

One of the most important steps in allocation decision is to set the position of the page according to page arrival time. The

allocation (placing) decision in this algorithm is based on the equation below:

(POSITION – B*R)%B) *[3], where B is a Number of Pages in the Block, and R is a Number of the Erased Bocks.*

Page arrival time refer to the time of the page when had been appeared in page access pattern. FCFS, initializing the block list then it set the position for the incoming pages according to their arrival time. Because FCFS places pages according to their arrival time, a data page with lowest arrival time value will be placed at first, and then the page with greater arrival time value will be the next page and so on. However, if the page appears another time in the same block, then FCFS algorithm will mark the old page as invalid one, the system in this case recognize the last page as the valid one [3].

B. First Re-Arrival First Serve (FRFS)

First Re-arrival First Serve (FRFS) [3] allocation algorithm had been emerged from previous FCFS allocation algorithm. Instead of placing data pages according to their arrival time property, FRFS algorithm depends on calculating the re-arrive time value for each incoming file data page in order to determine the correct position (location) for it. Re arrival time is the time when the page re-arrives. FRFS allocation algorithm, uses the minimal number of blocks, due to it reuses block as soon as possible. FRFS algorithm has three main allocation stages, in first stage, it computes the re-arrival time of each page by scanning through the entire access pattern and then in the second stage, it allocates a cell for each data page. The page having the earliest re-arrival time is assigned (0) value, the page with the second earliest re-arrival time is assigned (1) value, and so on. In the third stage it places the pages into the blocks according to the ordinal number they are assigned from the second stage. If this algorithm determines that a block had become in inactive state, it will erase the block and moving it to the end of the block list data structure so it can be reused later. FRFS keeps track of number of active blocks, and at the end, the maximum of these active block numbers is the number of blocks required by it [3]. The following presents the pseudo code for First Re-arrival First Serve allocation (FRFS) algorithm:

1. *Initialize the block list*
2. *Compute the re-arrival time for each block*
3. *Compute the order according to the re-arrival time*
4. *For each page in the page access pattern*
5. *Set the position of the page according to the order of the page*
6. *Place the page into the ((position-B*R)%B)th cell of the ((position-R*B)/B)th block in the block list*
7. *If the page appeared before*
8. *mark the cell it resided invalid*
9. *If the block where the page resided before is now inactive*
10. *move it to the end of the block list to be reused*
11. *count the number of the active block in the block list*
12. *compare it with the one from the previous iteration*

Algorithm 2. The Pseudo Code for First Re-arrival First Serve (FRFS) Allocation

Regarding to allocation procedure, First Re-arrival First Serve (FRFS) algorithm start by initializing the block list, and then it executes calculation process to compute the re-arrival time for each page. The placing pages procedure is based on the following equation:

(POSITION – B*R)%B) *[3], where B: Number of Pages in the Block, R: Number of the Erased Bocks.*

C. Online First Re-Arrival First Serve (On-Line Frfs)

Online First Re-arrival First Serve [3] allocation algorithm should analyze the entire page access pattern in order to make the allocation decisions. It takes the allocation decision as soon as the page request arrives. Two essential data structure are used in this algorithm, block list data structure, and the second one is a prediction table that contains prediction information for all the pages that have appeared. The block list data structure contains all the blocks of the memory and the blocks there are sorted by their identification numbers [3]. The prediction table data structure contains information for all pages that have been appeared. To access the prediction information faster, the prediction table should be placed in high speed processing time memory such as Random Access Memory (RAM). According to what previously mentioned, each page has two data type, the estimated arrival interval for the page and last time that the page appeared. When new data page appears, On-Line FRFS algorithm estimates arrival interval and set it to the default value, the default value, in this case, refers to the mean value of the intervals of all pages that had been observed in the past, then On-Line FRFS the last arrival time of the page to current time. After that, On-Line FRFS insert this entry into the prediction table. Then On-Line FRFS will update the estimated interval value and the length of interval between the current time and the previously arrival time of the page [10].

Mathematically, the (On-Line FRFS) algorithm computes the interval estimated length using the following equation:

New estimated arrival interval = ***R*ER + (1-R)*(T-LT)*** *[3], where: R is a Constant between (0) and (1), ER is the Old Estimated Interval of the Data Page, T is the Current Time, and LT is the Data Page Last Arrival Time.*

Many mathematical steps should be followed in order to determine the correct position for incoming data pages, the following pseudo code shows the basic allocation steps that data pages should be passed through during allocation process:

1. *compute the re-arrival time for each page*
2. *compute an order of re-arrival time*
3. *for each page in page access pattern*
4. *let K be the ordinal number*
5. *place the page into the K/Bth block in the block list*
6. *if the same page appeared before*
7. *mark the cell it resided invalid*
8. *if the block the page previously resides became inactive*
9. *erase and move it to the end of the block list*

Algorithm 3. Pseudo Code of On-Line First Re-arrival First Serve Allocation Algorithm

D. Probability Based Popularity Allocation Scheme

Probability based popularity [22] allocation algorithm consider as one of modern allocation algorithm. The allocation decision is based on the probability of the page which basically depend on the popularity of each unique page in the pages access pattern. Regarding to this algorithm, the distribution of pages in the page access pattern is assumed to follows the (ZIPFs Law) [22], where pages are stored in an increasing order, from the heights rank to the lowest rank. This distribution is based on pages frequency of the occurrences in the access pattern [22].

Depending on the number of times that each page is appearing into page access pattern, this algorithm divides the blocks into two categories HOT and COLD blocks. Probability based popularity allocation algorithm consist of two main components, access screening and allocation algorithm. Access screening percolate the type of data access to two categorization, read and write, the allocation algorithm part perform a distribution process of the access data. This component consist of two sub component, popularity interpreter and allocation engine.

Two main functions are performed by these sub components. Popularity interpreter determines the state of accessed data, either hot or cold. The allocation engine performs block selection process [22]. The distribution process performed by the allocation algorithm component will cluster the distribution into three main groups, maximum probability, median probability and minimum probability [22].

In this state, the pages that have probability value greater than or equal to median probability are classified as hot page, others page will be classified to cold pages. Probability Based allocation algorithm, divide the blocks of flash-based storage media into two categories, hot and cold blocks. The probability based popularity algorithm assumes that the number of hot blocks should be more that number of cold blocks. The amount of blocks which are subjected to be hot or cold is calculated using the following two equations:

1. $H = (MAX+MEDIAN)*B$.
2. $C = 1- H$ *[22]*.

Where, H is the Amount of HOT blocks, MAX is the maximum probability, MEDIAN is the median probability, B is the cod block position and C is the Amount of COLD blocks.

Regarding to above mentioned equations, The Probability based popularity allocation algorithm keeps each block not to be in active state in the whole allocation process. If such blocks are active during the whole allocation process then number of active blocks will be increased. The overall allocation algorithm is simplified as follow:

```
1.  get the probability of each page
2.  set the highest probability to MAX
3.  set the lowest probability to MIN
4.  get the middle probability and set the MEDIAN
5.  for each page in the access pattern
6.  get the page weight
7.  if weight >= MEDIAN
8.  set page as HOT
9.  write page into free pages within blocks [0,H 1]
10. else
11.  Set page as COLD
12. Write page into free pages within block [H,B 1]
13. count ACTIVE blocks
```

Algorithm 4. Pseudo Code for Probability Based Allocation Algorithm

E. Best-M Algorithm

Best-M allocation algorithm is stand on Block-Based Allocation concept. Simply it places all the requests for the same pages into the same block. As a result, each block contains only a single most up-to-date content of a page plus all the previous contents that all have been marked invalid. The idea of this algorithm is to assign page appearances to cells according to their difference. A difference between two page appearances is defined as the sum of the difference of their arrival time and the difference of their re-arrival time. The reason that it uses difference to allocate cells is that it is likely that all cells in the same block will be set to valid and invalid at about at the same time [19]. Best-M Algorithm could be simplified by the following pseudo code:

```
1. compute block index for page appearance
2.  initialize block list
3. search block list for block(i)
4. If block(i) not found in the block list
5. insert the new block in the block list
6.  set index(i) for this block
7. Place page into first cell of block(i)
8. If block became inactive
9. delete block index from block list(i)
```

Algorithm 5. The Pseudo Code for Best-M Algorithm.

Best-M allocation algorithm uses block list data structure to store blocks in it. The block in Best-M allocation algorithm has two properties, block index and number of cells for each block. When page arrive, Best-M algorithm, compute block index (i) for this page, then it search block list to find this block. If the requested block not found in the block list, Best-M insert new block in the block list and assign index (i) to it. After that, Best-M places the page into that block. However, block recycle process is invoked when the state of block became Inactive [11].

V. COMPARISION BETWEEN FLASH-BASED SSD ALLOCATION ALGORITHMS

A group of algorithms had been discussed and various allocation techniques were adopted by each previous mentioned algorithm. Table1 show a general review for these algorithms then we will compare these algorithms theoretically in order to clarify and explain the main differences between it.

TABLE II. A GENERAL COMPARISON BETWEEN ALLOCATION ALGORITHMS USED IN SSD STORAGE SYSTEMS.

Comparison factor	Algorithm				
	FCFS	***FRFS***	***Online FCFS***	***Probability base***	***Best_ M***
Algorithm Type	On Line	Off Line	On Line	On Line	Off Line
Allocation mode	Arrival time	Re-arrival time	Pages Arrival Interval	Page probability	Difference value
Futures Discriminated Degree	Sequential allocation style	Emerged from FCFS	Mathematical based allocation process	Divide blocks into HOT & COLD	Each block with same data pages

As shown in table 2, there are five main algorithms. Online and Offline are two main types of allocation scenarios. The Online type referring to online allocation mode and Offline type is the internal-style data real architecture style. The allocation mode factor explains the concept that the allocation procedure is based on. Generally, future discriminated degree referring to the basic concept(s) that were adopted to perform the allocation procedure. We had executed these algorithms in real environment (simulation environment). The results presented in table 3, showing the complexity of each algorithm. The complexity consists of two main parameters, space and execution time. The space criterion is referring to the number of cells that had been requested by the storage system when (n) number of incoming pages had to be written. The execution time referring to the time required to execute the specified (pre-defined) number of write operations that are waiting in the page access pattern queue.

TABLE III. EMPIRICAL COMPARISON BETWEEN ALLOCATION ALGORITHMS USED IN SSD STORAGE SYSTEMS.

Algorithm	Complexity		Proposed work
	Space required	***Execution time***	
First Come First Serve	O (nn)	O (mn)	allocation algorithm using arrival time parameter
First Re-arrival First Serve	O (nn)	O (mn)	allocation algorithm using re-arrival time parameter
On-line FRFS	O (n)	O (n3)	On-line allocation algorithm
Probability-based	O (mn)	O (n2)	allocation algorithm based on page popularity probability
Best Match (Best-M)	O (n)	O (n2)	block-based allocation algorithm using page appearance and cell difference to parameters

As mentioned in table 3, the probability-based, and Best Match (Best-M) allocation algorithms, had satisfied best degree comparing with other allocation algorithms, despite that fact that First Come First Serve, and First Re-arrival First Serve (FRFS), had satisfied less execution time comparing with other algorithms.

VI. CONCLUSION

The paper provided an overview of the available literature in data allocation techniques for flash-based SSD storage systems. We have introduced, analyzed and compared several allocation techniques. We highlighted the importance of allocation schemes in both reducing waste-space, and adjust the random-based style of data writing operations. Respectively, the findings were objectively reported and it may considered as researching reference for obtaining obviously documented knowledge, on many flash-based allocation strategies. Generally, as a conclusion, the storage devices with built-in allocation algorithm, performs better than those haven't one. More deeply future work on this type of allocation strategies is required.

ACKNOWLEDGMENT

This work was supported by a grant from the Universiti Sains Malaysia. We thanks and recognition go to my advisor, Associate Professor. Dr. Putra Sumari, who for helping us in this paper. Last but not least, the authors would like to thank the School of Computer Science, Universiti Sains Malaysia (USM) for supporting this study.

REFERENCES

[1] [1] S. Boboila, and P. Desnoyers, "Write Endurance in Flash Drives: Measurements and Analysis," in FAST'10 Proceedings of the 8th USENIX conference on File and storage technologies, USA, 2010.

[2] W. Bux, and I. Iliadis, "Performance of greedy garbage collection in flash-based solid-state drives," Performance Evaluation, Vol. 67, No. 11, pp. 1172-1186, 2010.

[3] L.-F. Chou, and P. Liu, "Efficient allocation algorithms for flash file systems, Parallel and Distributed Systems " in 11th International Conference on Parallel and Distributed Systems, Vol.1, No.1, pp. 634 - 641, 2005.

[4] S. Choudhuri, and T. Givargis, "Performance improvement of block based NAND flash translation layer," in Hardware/Software Codesign and System Synthesis, 5th IEEE/ACM/IFIP International Conference on, 2007, Vol. 5, No.2, pp. 257 - 262.

[5] M. A. H. Chowdhur, and K.-H. Kimy, "A Survey of Flash Memory Design and Implementation of Database in Flash Memory," in 3rd International Conference on Intelligent System and Knowledge Engineering, South Korea, pp. Vol.1, pp.1256 - 1259, 2008.

[6] P. Desnoyers, "Empirical Evaluation of NAND Flash Memory Performance," ACM SIGOPS Operating Systems Review, Vol. 44, No. 1, 2010.

[7] E. GAL, and S. TOLEDO, "Algorithms and Data Structures for Flash Memories," ACM Computing Surveys (CSUR), Vol. TBD ,No. TBD, pp. 1-30, 2005.

[8] X.-y. Hu, E. Eleftheriou, R. Haas et al., "Write Amplification Analysis in Flash-Based Solid State Drives," in The Israeli Experimental Systems Conference-SYSTOR 2009, Israeli 2009.

[9] H. H. Huang, S. Li, A. Szalay et al., "PerformanceModeling and Analysis of Flash-based Storage Devices," in 27th Symposium on Mass Storage Systems and Technologies (MSST), USA, 2011, pp. 1-11.

[10] M. Huang, O. Serres, V. K. Narayana et al., "Efficient cache design for solid-state drives," in Proceedings of the 7th ACM international conference on Computing frontiers, Bertinoro, Italy, 2010, pp. 356.

[11] Hyunchul Park, and D. Shin, “Buffer flush and address mapping scheme for flash memory solid-state drive ” Journal of Systems Architecture Vol. 56, pp. 208-220, 2010.

[12] A. Jiang, and J. Bruck, “Information Representation and Coding for Flash Memories,” in IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, USA,Vol. 6, No2 , pp. 920-925, 2009

[13] A. Kawaguchi, S. Nishioka, and H. Motoda, “A Flash-Memory Based File System,” in the Winter 1995 USENIX Technical Conference, USA, 1995, pp. 155-164.

[14] J. KIM, J. SEOL, and S. MAENG, “A Buffer Management Issue in Designing SSDs for LFSs,” TRANS.INF&SYST., Vol. E93-D, No. 6, pp. 1644-1647, 2010.

[15] Y. Kim, B. Tauras, A. Gupta et al., “FlashSim: A Simulator for NAND Flash-based Solid-State Drives,” in Proceedings of the IARIA International Conference on Advances in System Simulation (SIMUL), Portugal, 2009.

[16] C. King, and T. Vidas, "Empirical analysis of solid state disk data retention when used with contemporary operating systems," Digital Investigation,Vol, 8, pp. S111-S117, 2011.

[17] O. Kwon, J. Lee, and K. Koh, “EF-Greedy: A Novel Garbage Collection Policy for Flash Memory Based Embedded Systems ” in International Conference on Computational Science (4), China, Vol.4, No,2, pp. 913-920.

[18] S. J. Kwon, and T.-S. Chung, “An Efficient and Advanced Space-management Technique for Flash Memory using Reallocation Blocks,” in IEEE Transactions on Consumer Electronics,Vol.54, No.2, pp. 631 - 638

[19] P. Liu, C.-H. Chuang, and J.-J. Wu2, "Block-Based Allocation Algorithms for FLASH Memory in Embedded Systems," PARALLEL COMPUTING TECHNOLOGIES, pp. 569-578, 2007.

[20] E. Otoo, A. Pinar, D. Rotem et al., "A File Allocation Strategy for Energy-Efficient Disk Storage Systems," http://escholarship.org/uc/item/18p8g9z0, [28Nov2011, 2008].

[21] A. R. Rahiman, and P. Sumari, "Block Cleaning Process in Flash Memory," http://www.intechopen.com/articles/show/title/block-cleaning-process-in-flash-memory, [28 NOV 2011, 2011].

[22] A. R. A. Rahiman, and P. Sumari, “Probability Based Page Data Allocation Scheme in Flash Memory,” in ADVANCES IN MULTIMEDIA INFORMATION PROCESSING - PCM 2009, Thailand, 2009, pp. 300-310.

AUTHORS PROFILE

Jaafer Mohammad Al-Sabateen is a researcher in University Science Malaysia, School of Computer Science. He obtained software engineering Bachelor degree from Isra private university in 2008 and continued his study to a master degree in Computer Science.

Saleh Ali Alomari obtained his Bachelor degree in Computer Science from Jerash University, Jordan in 2005 and Master degree in Computer Science from Universiti Sains Malaysia (USM), Pulau Penang, Malaysia in 2007. Currently, He is a Ph. D. candidate at the School of Computer Science, Universiti Sains Malaysia. He is the candidate of the Multimedia Computing Research Group, School of Computer Science, USM. He is managing director of ICT Technology and Research and Development Division (R&D) in D&D Professional Consulting Company. He has published over 35 papers in international journals and refereed conferences. He is a member and reviewer of several international journals and conferences (IEICE, ACM, KSII, JDCTA, IEEE, IACSIT, etc). His research interest are in area of Multimedia Networking, video communications system design, multimedia communication specifically on Video on Demand system, P2P Media Streaming, MANETs, caching techniques and for advanced mobile broadcasting networks as well.

Putra Sumari obtained his MSc and PhD in 1997 and 2000 from Liverpool University, England. Currently, he is a lecturer at the School of Computer Science, Universiti Sains Malaysia, Penang. He is the head of the Multimedia Computing Research Group, School of Computer Science, USM. Member of ACM and IEEE, Program Committee and reviewer of several International Conference on Information and Communication Technology (ICT), Committee of Malaysian ISO Standard Working Group on Software Engineering Practice, Chairman of Industrial Training Program School of Computer Science USM, Advisor of Master in Multimedia Education Program, UPSI, Perak.

Building Trust In Cloud Using Public Key Infrastructure

A step towards cloud trust

Ms. Heena Kharche
Computer Science and Engineering
IES IPS Academy
Indore India

Mr. Deepak Singh Chouhan
Computer Science and Engineering
IES IPS Academy
Indore India

***Abstract*—Cloud services have grown very quickly over the past couple of years, giving consumers and companies the chance to put services, resources and infrastructures in the hands of a provider. There are big security concerns when using cloud services. With the emergence of cloud computing, Public Key Infrastructure (PKI) technology has undergone a renaissance, enabling computer to computer communications. This study describes use of PKI in cloud computing and provides insights into some of the challenges which cloud-based PKI systems face.**

Keywords- Cloud Computing; Public Key infrastructure; Cryptography.

I. INTRODUCTION

Cloud computing is rapidly emerging as a new paradigm for delivering computing as a utility. It allows leasing of IT capabilities whether they are infrastructure, platform, or software applications as services on subscription oriented services in a pay-as-you-go model.

"Cloud computing" is the next natural step in the evolution of on-demand information technology services and products. To a large extent, cloud computing will be based on virtualized resources.

Cloud Computing is here to stay, as it is proposed to transform the way IT is deployed and managed, promising reduced implementation, maintenance costs and complexity, while accelerating innovation, providing faster time to market, and providing the ability to scale high-performance applications and infrastructures on demand.

This paper will discuss why a cloud customer will trust cloud by using PKI.

This paper is organized in the following sections:

In Section II, we give a background of the technologies to be used. We explore in more detail what are the basic fears from cloud customers to adopt the cloud computing.

In Section III we describe the problems of cloud based PKI.

In Section IV we present our vision, some broad strategies that might be used to mitigate some of the concerns outlined in Sections II and III. Finally in section V we will discuss future work followed by conclusion in section VI.

II. BACKGROUND

A. Cryptography

Now a day, the most effective method of securing the data is by using cryptographic techniques. Cryptography is the method of storing and transmitting data in form that only those it is intended for can read and process [1].Basic terms used in cryptography are;

1. The readable data is referred to as PLAINTEXT
2. The random and unreadable data is referred to as CIPHERTEXT.
3. Process of converting plaintext to cipher text is referred to as ENCRYPTION.
4. Reverse of encryption i.e. Process of converting cipher text to plaintext is known as DECRYPTION.
5. Set of rules dictating how to encrypt and decrypt data are referred to as ALGORITHM.

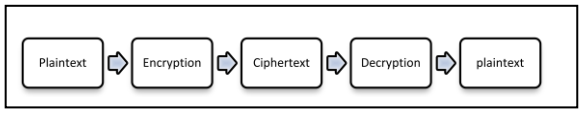


Figure 1. Cryptography process

1) Cryptosystem: The hardware or software implementation of cryptography process is termed as cryptosystem. Following services are provided by cryptosystems [2]:

1. **Confidentiality:** Confidentiality is the need to ensure that information is disclosed only to those who are authorized to view it.
2. **Integrity:** Integrity is the need to ensure that information has not been changed accidentally or deliberately, and that it is accurate and complete.
3. **Authentication:** Authentication is the process of confirming correctness of the claimed identity.
4. **Authorization:** Authorization is the approval, Permission or empowerment for someone to do something.

5. **Non repudiation:** Non Repudiation is the ability for a system to prove that a specific user and only that user sent a message and it hasn't been modified.

2) Public Key Cryptography: As discussed in RSA data security white paper [3] Cryptography uses mathematical algorithms and processes to convert intelligible plaintext into unintelligible ciphertext, and vice versa. Applications of cryptography include:

- Data encryption for confidentiality
- Digital signatures to provide non-repudiation (accountability) and verify data integrity
- Certificates for authenticating people, applications and services, and for access control (authorization)

The two main kinds of cryptography are shared secret (symmetric key encryption) and public key (Asymmetric key encryption).

3) Symmetric Key Encryption: In symmetric key encryption, encryption key can be calculated from the decryption key and vice versa. With most of the symmetric algorithms, the same key is used for encryption and decryption. The symmetric key is effective only when the key is kept secret by two parties if anyone else discovers the key in any way; it affects both Confidentiality and Authentication. A person with unauthorized symmetric key not only can decrypt messages sent with key but can encrypt new messages and send them on behalf of the legitimate parties using the key.

4) Asymmetric Key Encryption: Public key encryption also called as Asymmetric Encryption involves a pair of keys, a public key and a private key, associates with an entity. Each public key is published, and the corresponding private key is kept secret. Data encrypted with public key can be decrypted only with corresponding private key.

B. Public Key Infrastructure

PKI consists of programs, data formats, procedures, communication protocols, security policies and public key cryptographic mechanisms working in a comprehensive manner to enable a wide range of dispersed people to communicate in a secure and predictable fashion. PKI provides authentication, confidentiality, non-repudiation, and integrity of the messages exchanged. PKI is hybrid system of symmetric and asymmetric key algorithms and methods [1-3].

A public-key infrastructure (PKI) is a framework that provides security services to an organization using public-key cryptography. These services are generally implemented across a networked environment, work in conjunction with client-side software, and can be customized by the organization implementing them. An added bonus is that all security services are provided transparently— users do not need to know about public keys, private keys, certificates, or Certification Authorities in order to take advantage of the services provided by a PKI [4].

1) Components of PKI: As discussed in paper [5] there are five components in PKI:

a. End Entity: End Entity is a generic term used to denote end-users, devices (e.g., servers, routers), or any other entity that can be identified in the subject field of a public key certificate. End entities typically consume and/or support PKI-related services.

b. Certification Authority (CA): an entity which issues certificates. One or more in-house servers, or a trusted third party such as VeriSign or GTE, can provide the CA function

c. Registration Authority (RA): The RA is an optional component that can assume a number of administrative functions from the CA. The RA is often associated with the End Entity registration process, but can assist in a number of other areas as well.

d. Repository: A repository is a generic term used to denote any method for storing certificates and CRLs so that they can be retrieved by End Entities.

e. CRL Issuer: The CRL Issuer is an optional component that a CA can delegate to publish CRLs.

2) PKI and the Aims Of Secure Internet Communication:

The four aims of secure communication on the Internet are as stated earlier: confidentiality, integrity, authentication and non-repudiation. Authentication is the procedure to verify the identity of a user. There are three different factors authentication can be based on. These factors are something the user knows, something the user possesses and something the user is. Something the user knows could be a password that is a shared secret between the user and the verifying party. This is the weakest form of authentication since the password can be stolen through, for example, a dictionary attack or sniffing the network. Something the user possesses could be a physical token like a credit card, a passport or something digital and secret like a private key. This authentication form is usually combined with something the user knows to form a two-factor authentication. For instance, a credit card and a PIN are something possessed and something known. Something the user is could be something biometric like a fingerprint, DNA or a retinal scan which is unique for the user.

C. Cloud Computing

Cloud computing ('cloud') is an evolving term that describes the development of many existing technologies and approaches to computing into something different. Cloud separates application and information resources from the underlying infrastructure, and the mechanisms used to deliver those [6]. Cloud computing is defined as a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or cloud provider interaction [7].

1) Cloud services exhibit five essential characteristics [6-7] that demonstrate their relation to, and differences from, traditional computing approaches:

i. *On-demand self-service:* Computing capabilities are available on demand without any interference of third party.

ii. *Broad network access:* Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, laptops, and PDAs) as well as other traditional or cloud based software services.

iii. *Resource pooling:* The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. Examples of resources include storage, processing, memory, network bandwidth, and virtual machines. Even private clouds tend to pool resources between different parts of the same organization.

iv. *Rapid elasticity:* Capabilities can be rapidly and elastically provisioned, in some cases automatically to quickly scale out; and rapidly released to quickly scale in. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be purchased in any quantity at any time.

v. *Measured service:* Cloud systems automatically control and optimize resource usage by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, or active user accounts). Resource usage can be monitored, controlled, and reported providing transparency for both the provider and consumer of the service.

NIST Visual Model of Cloud Computing Definition gives the overall perspective and definition of what cloud computing is [8]:

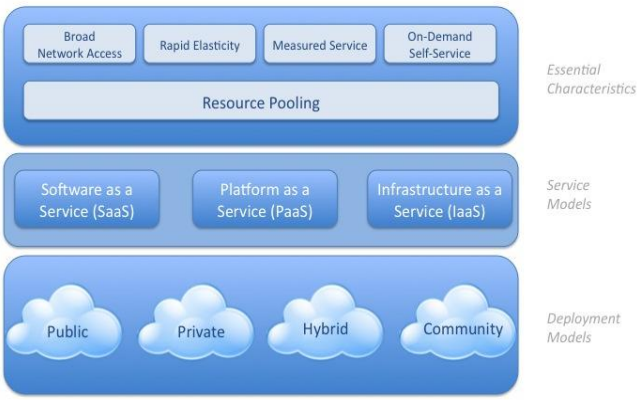


Figure 2. NIST Visual Model of Cloud Computing

a. Service Models: Cloud computing can be classified by the model of service it offers into one of three different groups.

i. **IaaS (Infrastructure as a Service):** The capability provided to the customer of IaaS is raw storage space, computing, or network resources with which the customer can run and execute an operating system, applications, or any software that they choose.

ii. **PaaS (Platform as a Service):** The cloud provider not only provides the hardware, but they also provide a toolkit and a number of supported programming languages to build higher level services (i.e. software applications that are made available as part of a specific platform).

iii. **SaaS (Software as a Service):** The SaaS customer is an end-user of complete applications running on a cloud infrastructure and offered on a platform on-demand. The applications are typically accessible through a thin client interface, such as a web browser.

b. Deployment Models: Clouds can also be classified based upon the underlying infrastructure deployment model as Public, Private, Community, or Hybrid clouds.

i. Public Cloud: A public cloud's physical infrastructure is owned by a cloud service provider. Such a cloud runs applications from different customers who share this infrastructure and pay for their resource utilization on a utility computing basis.

ii. Private Cloud: A pure private cloud is built for the exclusive use of one customer, who owns and fully controls this cloud.

iii. Community Cloud: When several customers have similar requirements, they can share an infrastructure and might share the configuration and management of the cloud. This management might be done by themselves or by third parties.

iv. Hybrid Cloud: Any composition of clouds, be they private or public could form a hybrid cloud and be manage a single entity, provided that there is sufficient commonality between the standards used by the constituent clouds.

2) Complications in cloud:

There are some concerns that should not be taken lightly when moving to a cloud service. Once the data has been moved to a cloud provider, control over it has been lost. The user cannot tell where the data resides physically and cannot be fully confident the data is handled with care in a secure manner. Furthermore, when the data has been moved to or created in the cloud, there are concerns about who really owns the data. For instance, if the subscription to the cloud service is cancelled, the customer cannot be fully confidant the data is removed. Also, if the customer wishes to switch cloud provider, there are concerns about if it is even possible as the provider might lock in the customer with various methods. Providers are very much aware of the complications and concerns of their services and works constantly to improve the quality and security of their services. Among others things they usually employ IT-security staff 24 hours a day every day to cope with any upcoming problems.

Benefits of cloud Computing: According to Mike Klein [12] there are six strong benefits that a cloud user gets:

i. *Lower Costs:* Cloud computing pools all of the computing resources that can be distributed to applications as needed – optimizing the use of the sum of the computing resources and delivering better efficiency and utilization of the entire shared infrastructure.

ii. *Cap-Ex Free Computing:* Whether you go with a public cloud or outsourced private cloud computing option, cloud computing delivers a better cash flow by eliminating the capital expense associated with building the server infrastructure.

iii. *Deploy Projects Faster:* Because servers can be brought up & destroyed in a matter of minutes, the time to deploy a new application drops dramatically with cloud computing. Rather than installing and networking a new hardware server, the new server can be dialed up and imaged in through a self-serve control console.

iv. *Scale as Needed:* As your applications grow, you can add storage, RAM and CPU capacity as needed. This means you can buy "just enough" and scale as the application demands grow. This benefit includes elasticity of the resources.

v. *Lower Maintenance Costs***:** Driven by 2 factors: Less hardware and outsourced, shared IT staff. Because cloud computing uses less physical resources, there is less hardware to power and maintain. With an outsourced cloud, you don't need to keep server, storage, network, and virtualization experts on staff full time.

vi. *Resiliency and Redundancy:* One of the benefits of a private cloud deployment is that you can get automatic failover between hardware platforms and disaster recovery services to bring up your server set in a separate data center should your primary data center experience an outage.

D. Transport layer Security:

Transport Layer Security (TLS) is a protocol that ensures privacy between communicating applications and their users on the Internet. When a server and client communicate, TLS ensures that no third party may eavesdrop or tamper with any message. TLS is the successor to the Secure Sockets Layer (SSL) [9].

Transport Layer Security (TLS) and its predecessor, Secure Sockets Layer (SSL), are cryptographic protocols that provide endpoint authentication and secure communications over any transport. TLS is normally associated with Internet communication but can be applied to any transport layer, including sockets and HTTP. TLS allows for two levels of security: Server Authentication and Mutual Authentication [10].

Server Authentication: Server Authentication authenticates the server to the client. When server authentication is used, the end user, or client, verifies that the server they are communicating with is actually who it says that it is. In the Internet world, your browser is the client, and a website such as Amazon™ is the server. Millions of clients need to be able to prove that the site to which they are giving financial information is really Amazon.

Mutual Authentication: Mutual Authentication authenticates the server to the client, and the client to the server. When Mutual Authentication is used, both the client and the server provide and validate certificates in order to verify each other's identity.

The TLS protocol is made up of two layers [11].

- The TLS record protocol is designed to protect confidentiality by using symmetric data encryption.
- The TLS handshake protocol allows authentication between the server and client and the negotiation of an encryption algorithm and cryptographic keys before the application protocol transmits or receives any data.

Need of TLS: Sending unencrypted messages increases the risk that messages can be intercepted or altered. TLS security technology automatically encrypts e-mail messages between servers thereby reducing the risk of eavesdropping, interception, and alteration.

III. ISSUES IN CLOUD BASED PKI

According to us there can be three issues that can complicate the implementation of PKI on cloud:

A. Storing Private Keys In Scalable And Mobile Systems:

The three factors to consider when designing the system are scalability, mobility and automation. A solution must be able to add more CAs on demand, be relatively consistent in required time to sign certificates and always be available. Hence, the solution must support the CA operations being movable to another less strained server if the number of requested signatures increases beyond the limit of the Hardware Security Module or the service unexpectedly fails. To able to move all CA operations to another server, all data regarding that CA must be moved between databases and the private key has to be moved or be the same at the new location. However, there exists no sufficiently secure procedure to move private keys between HSMs autonomously. Therefore, the same private keys must be predefined in HSMs at all available locations of that CA. The ability to move the CA to another location and to bind private keys on demand provides scalability in the number of signatures the system can handle. The scalability of the number CAs at one location is relative to the number of keys the Hardware Security Module is able to store.

B. Certificate Authority Separation:

One essential requirement of a cloud based PKI is that one customer should only be able to see and use its own CAs. Consequently, there must be separation between CAs and customers.

C. Providing Secure Authentication And Authorization:

Only a number of predefined CAs can issue certificates to administrators due to the trust store in the application server. Other CAs issuing administrator certificates can be added but that requires restarting of the application server. The purpose of this is to give each customer a dedicated CA to issue certificates to its administrators.

IV. Proposed Solution

In order to build complete trust over CA we must use the model suggested in RSA Conference Europe 2011 [12]. Studying the model we can establish trust of cloud consumer in cloud.

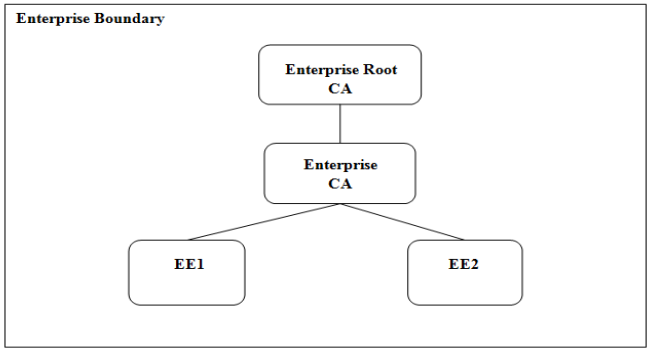


Figure 3. Enterprise root CA and Enterprise CA

An Enterprise Certificate Authority (CA) is a CA which generates certificates for a restricted community such as for an organization. The enterprise CA is selected by the mutual understanding of the enterprises and participating enterprises should trust the enterprise CA and Enterprise root CA.

The above model has following advantages:

1) Trust points (Root CA and CA certificates) are inside the enterprise boundary: All the enterprisess now can trust the root CA and CA as they are inside their boundary and the data kept on the cloud will not pass through the enterprise boundary.

2) Full control of security properties of PKI: In this model we will get the full functionalities of the Public Key Infrastructure that is the management of keys and the trust boundary will not be beyond the enterprise boundary.

3) On demand certificate and certificate revocation issuing: As both the enterprise CA and root CA are in the enterprise boundary any enterprise can easily revocate or issue a new certificate without have to wait for a long procedure and time.

The limitation of the mentioned model is, only few browsers will support the certificates issued by Enterprise root CA. The above limitation can be ignored as enterprise group can use a common browser to interact with each other.

Following aspects are considered while using the system:

a. Browser to be used: As the root CA is the enterprise CA therefore, the Certificates issued by the root CA will be supported only by few browsers. In order to interact with the cloud the enterprise should be aware of which browser is to be used.

b. Selection of Enterprise root CA and CA: Both the enterprise root CA and CA must be selected mutually by all the enterprises. A set of rules must be abided by the enterprises while choosing the and CA.

c. Switching cloud outside enterprise boundary: If any of the enterprise, at any time wants to switch the data from the enterprise boundary to outside world it can do so by mutually signed agreements and set of rules.

V. Future Work

A. Cross-Certification And Building Additional Certification Paths Dynamically

As it is a cloud service, it has the ability to cross certify the email authentication CAs from different customers under one central root email CA. In this scenario every customer that satisfies the criteria of the email certification will be offered to have the email CA certified and join the group of trusted email CAs. The implication would be that all customers in a group could send secure emails to each other and have access to the certificates and revocation information. The scenario relies on dynamically adding the certification path without having to reissue the certificates, which may or may not be possible.

B. Caching

Online Certificate Status Protocol (OCSP) is a very simple request/reply protocol that allows clients to ask an "OCSP responder" about the revocation status of one or more certificates. The OCSP responder returns digitally signed responses regarding the status of the certificates identified in the request. OCSP is designed to return real time responses to client queries, and can provide an efficient method for returning certificate status on demand.

To minimize the workload of the OCSP responder, caching of the result for some time could be used. However, OCSP use POST to send data which should not, in contrary to GET, be cached according to RFC 2616. Hence, further studies in the field of caching an OCSP response are required.

VI. Conclusion

PKI is enabling computer to computer communications in The Cloud because it offers a cryptographically strong method of authentication which can be tied to the secure transport mechanism, TLS [12].

The security of any system is not a question of if the system is secure or not, it is a question of how secure it is or in other words, to what extent it is secure. Every system has flaws, either in the design or in the nature of the system, thus absolute security cannot be guaranteed for any system. Technologies and incentives to access or destroy systems emerge as technology moves forward and the value of the system increases. Hence, a system can only be classified secure to an extent or not secure at all.

One critical factor in security is cost. To limit the incentives to break the system, the cost of breaking the system should be higher or equal to the value of the information the system is protecting. The paper has discussed a model to build trust in Cloud using public key Infrastructure. Despite of the limitation of browser support it can be widely used by

enterprises. The application of the above model can be the different plants and branch offices that want to share same data can create a public cloud and define their own Root CA and CA to ensure confidentiality and integrity of the data.

While working on future scope we can easily make a cloud consumer trust that their data is safe on cloud that too within their own enterprise boundary.

REFERENCES

[1] Shon Harris, CISSP All-in-One Exam Guide, Fifth Edition.

[2] Glossary of terms in SANS reading room available at: http://www.sans.org/security-resources/glossary-of-terms/

[3] Understanding Public Key Infrastructure (PKI) An RSA Data Security White Paper available at: ftp://ftp.rsa.com/pub/pdfs/understanding_pki.pdf

[4] Article from Entrust.com available at: http://www.entrust.com/resources/pdf/whatsapki.pdf

[5] Shashi kiran, Patricia Lareau, Steve Lloyd PKI Basics - A Technical Perspective available at: http://www.oasis-pki.org/pdfs/PKI_Basics-A_technical_perspective.pdf

[6] Security Guidance for Critical Areas of Focus in Cloud Computing V2.1 available at: https://cloudsecurityalliance.org/csaguide.pdf

[7] Guidelines on Security and Privacy in Public Cloud Computing Wayne Jansen Timothy Grance Draft Special Publication 800-144 available at: http://csrc.nist.gov/publications/drafts/800-144/Draft-SP-800-144_cloud-computing.pdf

[8] cloud computing architectural framework in CSA found at https://wiki.cloudsecurityalliance.org/guidance/index.php/Cloud_Computing_Architectural_Framework

[9] Blog by Mikko Nieminen available at: http://searchsecurity.techtarget.com/definition/Transport-Layer-Security-TLS

[10] VMware vCenter Configuration Manager Transport Layer Security Implementation WHITE PAPER available at: http://www.vmware.com/files/pdf/techpaper/vcenter-configuration-manager-transport-security-layer-tls-guide.pdf

[11] Transport Layer FAQ available at: http://www.bnymellon.com/security/tlsencryption.pdf

[12] PKI reborn in cloud by Jaimee Brown and Peter Robinson RSA, The Security Division of EMC found at: http://365.rsaconference.com/servlet/JiveServlet/previewBody/3037-102-1-4074/NMS-301%20-%20PKI%20Reborn%20in%20the%20Cloud.pdf

AUTHORS PROFILE

Ms. Heena Kharche
Pursuing Masters of engineering 2nd year
Computer Science and Engineering
Institute of Engineering and science Indore professional Studies Academy Indore India
heenak.28@gmail.com

Mr. Deepak Singh Chouhan
Working in Computer Science and Engineering department,
Institute of Engineering and science Indore professional Studies Academy Indore India.
deepak.ur@gmail.com

Contextual Modelling of Collaboration System

Purchasing process application

Wafaa DACHRY [#1], Brahim AGHEZZAF [#1], Bahloul BENSASSI [#2], Adil SAYOUTI [#3]
[#1]Department of Mathematics and Computer Faculty of sciences, Hassan II University
[#2]Department of physics, Faculty of sciences, Hassan II University
[#3] Systems Architecture Team, ENSEM, Hassan II University
BP 5366, Maarif, Casablanca 20100, Morocco

***Abstract*— Faced with new environmental constraints, firms decide to collaborate in collective entities and adopt new patterns of behavior. So, this firms' collaboration becomes an unavoidable approach. Indeed, our aim interest in our study is to propose a collaborative information system for supply chain. Our proposed platform ensures cooperation and information sharing between partners in real time. In fact, several questions have to be asked: What is the information nature may be shared between partners? What processes are implemented between actors? What functional services are supported by the platform? In order to answer these questions, we present, in this article, our methodological approach of modelling, called CMCS (Contextual Modelling of Collaborative System).**

Keywords-collaborative information system; business process; collaborative process; BPMN; CMCS.

I. INTRODUCTION

The integration of a collaborative information system within the supply chain has become a key element for improving manufacturing performance [1]. The collaborative information system aims to ensure coordination between different actors in the supply chain. It consists of a set of services developed to better meet user requirements.

To this end, we propose a modelling approach which consists to ensure the evolution of information through the realization of a set of models (business and applicative). These models offer a global vision of the company, partners and collaborative systems. A main feature of our approach is flexibility. This property can be defined as the system ability that can help system to easily adapt to the requested evolutions. The flexibility notion is provided through the separation of the business view from the applicative and technical one. So, we model the business processes and we identify necessary services regardless of application characteristics and technics.

The methodology is based on process approach [2] and adapted to the urbanism model of information system [3].

The business view: it lists the business events that the company must treat, business processes that respond to these events, documents used in these processes and actors who execute them.

The functional view: it describes the functions of information system, such as they are described in the specifications for the implementation of a new application. Functions can be: the management of a customer's contract, the management of delivery, etc.

The application view: It lists all the applications used by actors to perform functions and equip processes.

The Physical view: It lists all the infrastructure components (hardware, network infrastructure, equipment security, archiving, etc.) that support the system.

So, we present an adaptive model to the urbanism model of information system [4] [5], by adding two other necessary views for the implementation of our collaborative system (strategy and service). Also, the business view contains two contexts (business and collaborative ones).

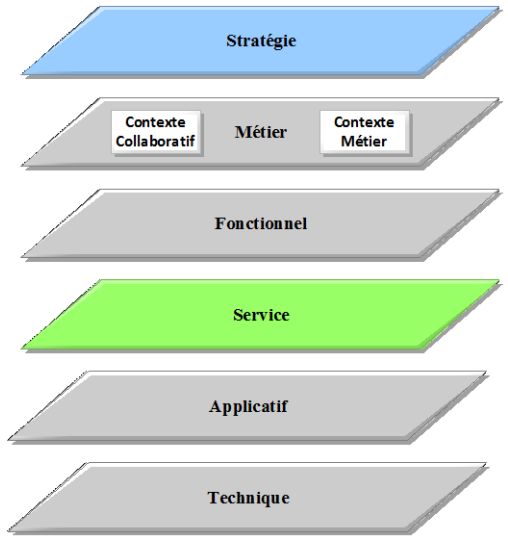


Figure 1. Adaptive model of urbanism model

The principle of CMCS can be summarized as follows: After defining the key processes used in the implementation of our system, they are detailed in activities to bring out the interactions between actors, modelling collaborative processes and determine various business services needed to carry out these processes.

The function blocks are then added in order to coherently organize all business objects. Functional services, which perform the function blocks defined in advance, are exposed by IT services (CRUD service and technical service [6]).

This paper describes different phases of CMCS approach (Contextual Modelling of Collaborative System) and a practical example for modelling the purchasing process

II. CONTEXTUAL MODELLING OF COLLABORATION SYSTEM

CMCS is a methodological approach in two phases (Study needs, Modelling and Design of the system), each phase contains one or more steps. We present in detail the different phases.

A. Phases of the methodology

Phase 1: The study needs

The study needs, first phase of our project, is one of the most difficult phases of our approach. The following phase of CMCS is based on the deliverables of the "study needs".

This phase includes two major steps: developing the business motivation model and the organization of the company.

Deliverable 1: BMM (Business Motivation Model);

Deliverable 2: Process mapping;

Step 1: Development of business models of motivation

As a preliminary to the establishment of a collaborative architecture, it is important to define objectives and motivations that lead to achieving them. This step should give birth to a business plan indicating the means to implement [7]. To determine these business motivations, we relied on the standard (Business Motivation Model - BMM) proposed by the OMG. It provides an organized structure for the specification and definition of business plans. The construction of business models and application models must be deducted from objectives and strategies, which justifies the use of this standard by our approach CMCS. In the context of CMCS, we focus on the two entities (means and ends) in the development of BMM.

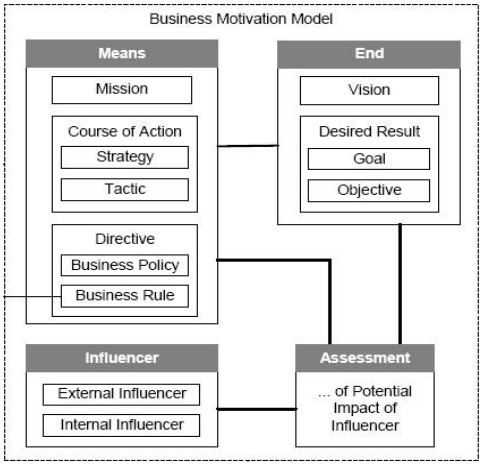


Figure 2. Business Model of Motivation [8]

Step 2: Study of business organization

This step of the approach follows the urbanization principle. One of the main phases of urbanization is the study of enterprise architecture. This involves identifying a set of mapping organizational of the company and its partners. The mapping is used to represent the environment already studied and how it works. This tool is essential at this step; however it must meet a few criteria:

- Global: The mapping should cover all business processes in order to have an overview of operations.
- Systemic: The mapping must be based on a systems approach [9]; the company should be considered as a set of micro-enterprises (subsystems) that operate with each other.
- Understandable: The mapping must be clear and understood by the entire company to allow many uses (business analysis, functional analysis, application analysis, etc.).

Phase 2: Modelling and Design of the system

The second phase of the process CMCS is the modelling and the design of system. This phase corresponds to the two views (business view and functional view) of the urbanization principle. The objective is to define the various functional necessary services to achieve information system. Various steps constitute this second phase:

Step 1: Business process modelling

Deliverable 1: Model the business process

Step 2: Modelling the collaborative business process

Deliverable 2: Model of the collaborative process

Step 3: Analysis and design of the system

Deliverable 3: UML diagrams (use case, sequence, class)

Step 4: Identification of functional services

Deliverable 4: Models of functional blocks and list of services

Step 1: Business process modelling

This step is based on the mapping of business processes. This step is used to describe business processes and all the activities constituting these processes.

The business processes must meet to one of objectives defined in the business motivation model of the company.

Step 2: Modelling the collaborative business process

To model collaborative processes, we rely on models of business processes of this phase first step. This model follows the BPMN formalism [10] [11], this concept is developed by the OMG; it's a communication tool that facilitates the rapid transition of the design of business processes to their implementation.

The objective of this step is the identification of business activities supported by the collaborative information system.

Step 3: Analysis and design of the system

This step consists to create an abstract representation of our system. It requires the use of an adequate method for the creation of different models on which we rely for the implementation of the system. The step aims to study

expectations of users and to identify elements involved in the collaborative system, their structures and their relationships.

In addition, the design consists to provide technical solutions to definite descriptions in the analysis. Object-oriented design focuses on defining software objects and how they work together. We use in this step the language UML (Unified Modelling Language) [12] in order to present three diagrams (use case, sequence, class).

Step 4: Identification of functional services

Based on the previous two steps, the description of collaborative business processes and analysis and design system, we generate the functional blocks [13] that include functional necessary services for the implementation of collaborative information system.

III. Application Of Our Approach CMCS

The case concerns a company specialist on manufacturing, distribution and installation of products and solutions in industrial automation and electrical engineering.

Faced with the pressures of competition and growth issues importance, the company is looking more and more for improving its industrial performance in terms of cost, time, adaptability and traceability. Indeed, it wants its business processes to be more flexible in order to support strategic developments of the company; such as participation in scenarios with other business. The company recognized the need of a stronger coordination and collaboration. So, it wants to develop a platform that allows it to communicate with partners and to exchange information.

The objective is the establishment of a collaborative information system corresponding to the company's business processes and to the collaborative ones of inter-companies network. In this context, we apply our approach to provide a flexible architecture that meets this company needs.

Phase 1: The study needs

Step 1: Development of business models of motivation

The first step is to develop the business motivation model. Remember that the fundamental concepts of the model are: Ends (include the vision, goals and objectives) and means (include the mission, strategies and tactics).

The vision describes the company's look in the future without worrying about how to achieve it. The goal expresses a state in order to meet this vision. The objective is a measurable component that we can achieve with some time constraint to meet the goal. As for the mission, it represents the operational activities of the company that can achieve the vision. Finally, Strategies represent actions needed to achieve goals.

By following these concepts, we develop an example of BMM.

Step 2: Study of business organization

The result of business organization leads us to develop a graphical representation of business processes.

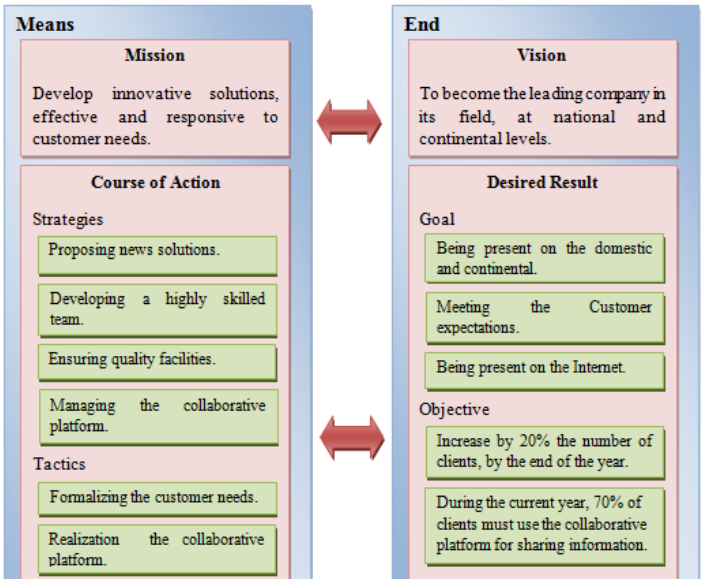


Figure 3. Example of BMM

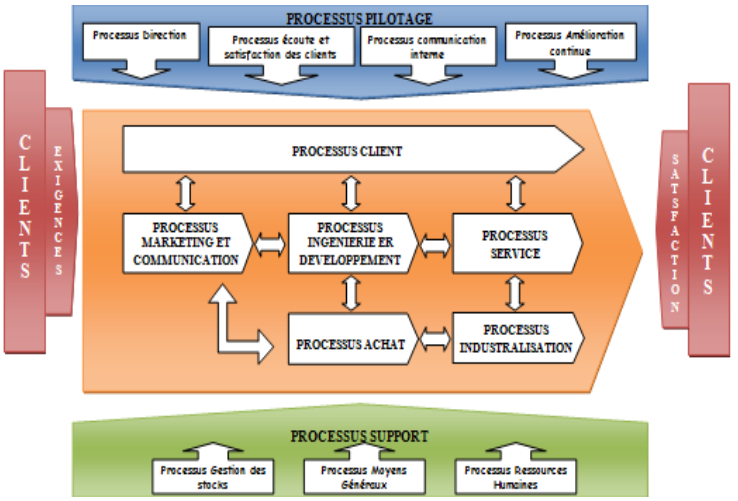


Figure 4. Process mapping

Phase 2: Modelling and Design of the system

Following our analysis, we focus on a specific process, we choose the purchasing one. Its main objective is to make goods or services available to internal users of the company with the best price, time and quality. Indeed, this process seeks to reduce the production costs, to ensure compliance of schedules and quality expected by customers.

Step 1: Business process modelling

We detail different activities of the purchasing process (see figure 5). The process begins with the definition of requirements formalized in a purchase request. This request generally specifies the following information:

- Description, type and characteristics of the product / service;
- The quantity to order;
- The delivery quality and desired time;

- The budget for the purchase.

The expression of needs is considered as an event triggering the purchasing process. This can be done by the various departments of the company.

- Analysis of the purchasing needs
- Research suppliers and request for proposals
- Analysis of proposals and negotiation
- Establishment of the order
- Receipt of products and quality control
- Invoice Verification

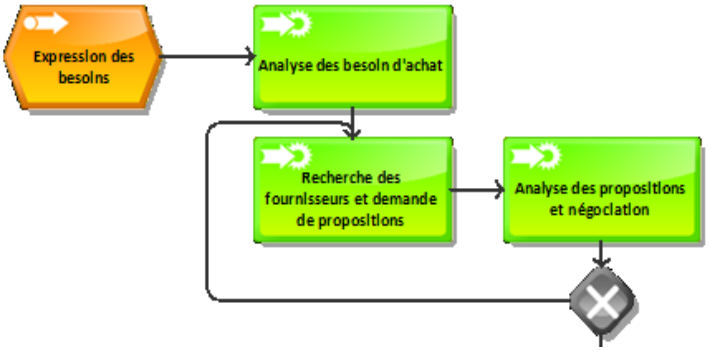


Figure 5. Part of the purchasing business process

Step 2: Modelling the collaborative business process

The purchasing collaborative process is modelled from the preceding description of business process. In this step, we describe the activities of partners involved in the collaborative process. In our case, collaborative alliances are already established between the producer (manufacturer) and partners of supply chain. The partnership created between the producer and supplier leads to the establishment of a framework-contract (legal support associated with partnership relations) between the two actors.

To study various acts of the purchasing collaborative process, we use the formalism BPMN (Business Process Modelling Notation). This notation will allow us to present the manual activities performed by actors of the chain (manufacturer and supplier of raw material) and automated activities supported by the collaborative platform.

In this context, we present two POOL partner (pool producer and pool supplier) and a POOL SIC. It contains two LANE (manufacturer module and supplier module). The figure below is an example of the purchasing collaborative process.

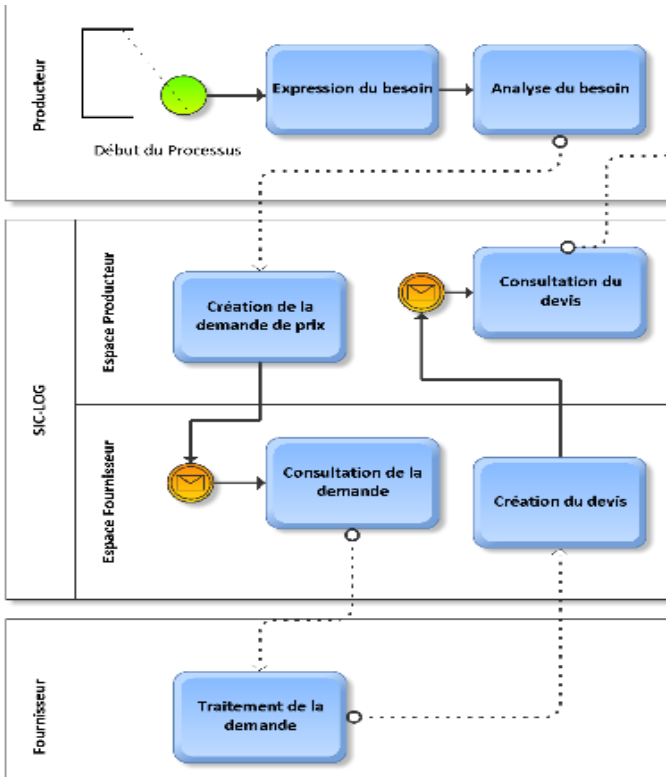


Figure 6. Part of collaborative process model

Step 3: Analysis and design of the system

- Diagram of use case

The use case diagram presented below shows few functionalities of managing the purchasing process.

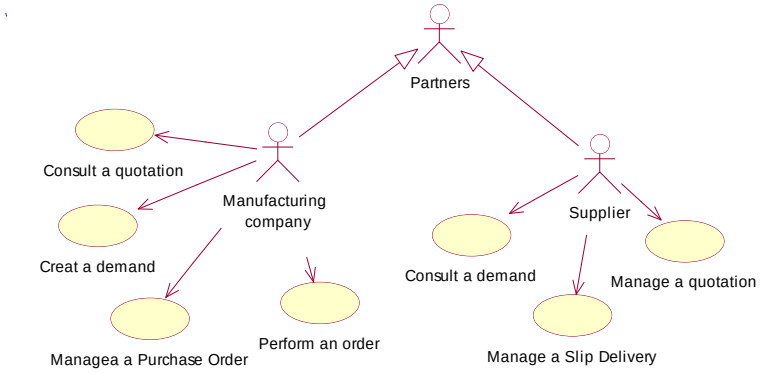


Figure 7. Use case diagram

- Sequence diagram

The sequence diagram shows different interactions between actors of the purchasing process.

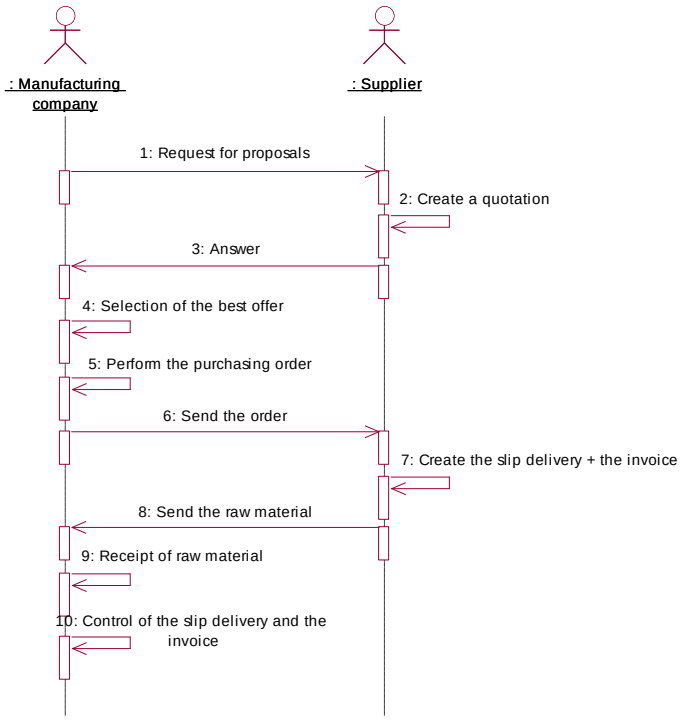


Figure 8. Sequence diagram

- Class diagram

The class diagram presented below expresses the static structure of the purchasing process.

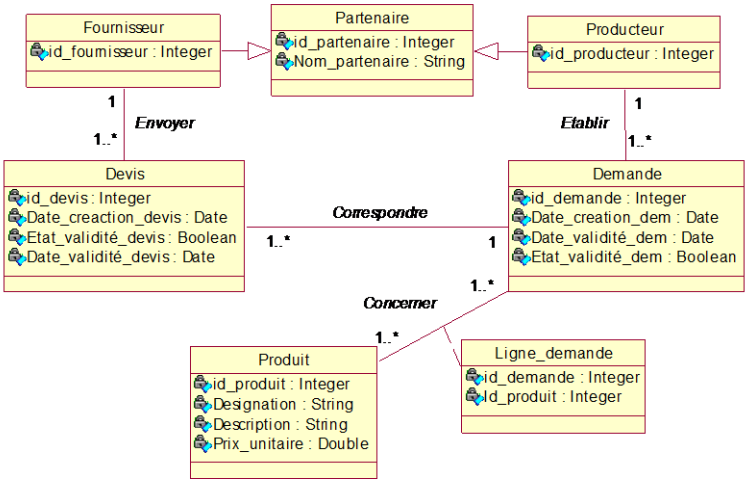


Figure 9. Class diagram

Step 4: Identification of functional services

We present in this step, functional blocks of our example. It represents functional services implemented in the system.

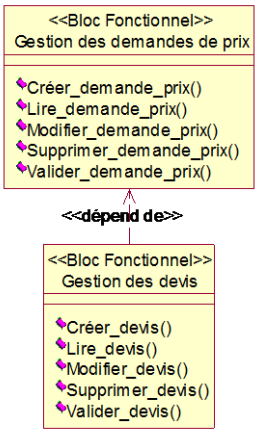


Figure 10. Functional Blocks

IV. Conclusion Et Perspectives

In this paper we presented the methodological approach CMCS (Contextual Modelling of Collaborative System) for modelling and design of the collaborative information system dedicated to manage the supply chain. We used formalism BPMN and UML for modelling. As perspective, we are proposing the architecture of the system based on multi-agent systems and we are developing an example of system.

Acknowledgment

I would like to thank to my advisors Mr. B.AGHEZZAF and Mr. B.BENSASSI, for their invaluable guidance and many useful suggestions during my work on this paper. I would also like to express my gratitude to all those who gave me the possibility to complete this paper.

References

[1] W. Dachry, B. Aghezzaf and B. Bensassi, “Integration of a collaborative information system in the global supply chain,” the first International Congress: Computing and Engineering Sciences, ISI 2011, Meknes - Errachidia, June 1 to 5, 2011

[2] H. Brandenburg, J-P. Wojtyna, “Process approach - Instructions for use,” Eyrolles, 2006

[3] Sqli Consulting, “SOA book, guide the architect of the SI,” 2nd edition Dunod, Paris, 2008

[4] K. Ben Driss, “SOA : pour une interopérabilité intrinsèque du Système d’Information,” Tozeur, Tunisie, Novembre 2008

[5] R. Courdier, “Système d’Information Urbanisme des SI,” http://personnel.univ-reunion.fr/courdier/cours/si/index.html, Date last accessed January 2012

[6] C. Devaux, “ Urbanization and architecture service oriented (SOA),” http://www.aubay.com/. 2008

[7] K. Boukadi, “Coopération interenteprises à la demande : Une approche flexible à base de services adaptables,” Thesis, November 2009

[8] Business Motivation Model, http://www.omg.org/spec/BMM/1.0/PDF, published in 2008

[9] A. Gautier, “Guide of implementation of the Systems Approach,” 6th European Congress of Systems Science September 2005

[10] http://www.bpmn.org/

[11] Business Process Modeling Notation (BPMN), Version 1.0, May 2004.

[12] P. Roques, "Uml in practice," Eyrolles, Mars 2008

[13] J. Simonin, " Functional design of enterprise services," RSTI - ISI pages 37 to 61, 2010

AUTHORS PROFILE

Wafaa Dachry was born in Casablanca in 1985. She received his professional master in logistics engineering, in 2008, from the faculty of science of Hassan II University, Casablanca, Morocco. In 2009, she joined the computer lab and decision support of the faculty of science, Hassan II University. Her actual main research interests concern collaborative information system for supply chain.

Brahim Aghezzaf is a professor in department of Mathematics and Computer, faculty of science, Hassan II University, Casablanca, Morocco. He is responsible of professional master in engineering and optimization of systems and logistics. Her research focuses on multiobjective optimization, on metaheuristics, on operational researchand constraint programming...etc.

Bahloul Bensassi is a professor in department of physics, faculty of science, Hassan II University, Casablanca, Morocco. He is responsible of professional master in logistics engineering and professional master in Electronics, Electrical, Automation and Industrial Informatics. Her actual main research interests concern logistics, electronics, automation and industrial informatics...etc.

Adil Sayouti received his Degree in High Education Deepened in physical science in 1999 from Hassan II University, Casablanca, Morocco .In 2001 he obtained his diploma of superior studies in computer science and in 2003 a Microsoft Engineering Systems Certificate. In 2005 he joined the system architecture team of the ENSEM, Casablanca, Morocco. He received her Ph.D. from in 2009. His actual main research interests concern Remote Control over Internet Based on Multi agents Systems.

Development of a Mobile Phone Based e-Health Monitoring Application

Duck Hee Lee
Department of Electrical Engineering
University of North Dakota
Grand Forks, North Dakota, USA

Ahmed Rabbi
Department of Electrical Engineering
University of North Dakota
Grand Forks, North Dakota, USA

Jaesoon Choi
Korea Artificial Organ Center, College of Medicine
Korea University
Seoul, South Korea

Reza Fazel-Rezai
Department of Electrical Engineering
University of North Dakota
Grand Forks, North Dakota, USA

***Abstract*—The use of Electrocardiogram (ECG) system is important in primary diagnosis and survival analysis of the heart diseases. Growing portable mobile technologies have provided possibilities for medical monitoring for human vital signs and allow patient move around freely. In this paper, a mobile health monitoring application program is described. This system consists of the following sub-systems: real-time signal receiver, ECG signal processing, signal display in mobile phone, and data management as well five user interface screens. We verified the signal feature detection using the MIT-BIH arrhythmia database. The detection algorithms were implemented in the mobile phone application program. This paper describes the application system that was developed and tested successfully.**

Keywords-Electrocardiogram(ECG); mobile phone; MIT-BIH database; health monitoring system.

I. INTRODUCTION

Nowadays, cardiac diseases are increasing in an alarming rate. According to the World Health Organization (WHO), cardiac disease is one of the leading causes of death in the developing world and is the leading cause in the developed world [1]. For these reasons, electrocardiogram (ECG) monitoring and diagnosis system is widely studied. ECG examination is a basic diagnosis procedure to find out if the patients have sporadic heart diseases, such as, arrhythmia and ischemia. Due to the growth of microcontroller and semiconductor technology, new ECG systems of small size and light weight have arrived [2]. Recent technological advances in wearable sensor networks, integrated circuits and wireless communication allow the design of light weight, low power consuming sensors at low-cost. Wearable and portable monitoring systems of physiological parameters have been studied by many research groups [3][4]. However, the majority of such health's monitoring devices are not suitable for medical monitoring of high-risk patients. Some of these systems have wireless modules, for instance Bluetooth and Radio Frequency (RF), but the development of local area networks (LAN) in hospitals have not matured yet. But a light and portable type wireless physiological signal-retrieving system has always been a medical personnel's dream [5]. A portable smart mobile phone has various functions. In medical fields, a new generation of mobile phones will have an important impact for the development of such healthcare systems, as they seamlessly integrate a wide variety of networks and thus provide the opportunity to transmit recorded biomedical data to a server in a hospital. Consequently, this paper describes the design and implementation of a prototype mobile healthcare application system and monitoring the ECG signals of patients in real-time.

II. METHOD AND MATERIAL

The system receiving block diagram is shown in Figure 1 of which the design and architecture details are explained in the following sub-sections.

A. Software Implementation

Development of software depends on operating system (OS) of mobile device. The emergence of various form of personal mobile device and associated various OS makes it important to make a smart choice based on the application requirements. We developed the portable monitoring system prototype using the SGH-i900 Omnia SmartPhone (Samsung Co. Ltd). This mobile device is equipped with a 624 MHz Marvell PXA312 processor, internal 16 GB storage and it includes a Bluetooth v2.0 interface. This mobile device supported the Windows Mobile 6.1 Professional for Marvell processor. Therefore, this mobile device is suitable for use in this research.

B. Signal Transmission Structure

Bluetooth is an industrial specification for wireless Personal Area Networks (PANs). Bluetooth provides a way to connect and exchange information between devices such as mobile phones, laptops, PCs, printers, digital cameras, and video game consoles over a secure, globally unlicensed short-range radio frequency. The Bluetooth specifications are developed and licensed by the Bluetooth Special Interest Group. It is a

standard communications protocol primarily designed for low power consumption, with a short range (1 meter (0 dBm), 10 meter (4 dBm), and 100 meter (20 dBm)) [6]. We used the small size (18*20*12 mm), low power consumption, high reliability, and low cost FB155BC (Firmtech Co., Ltd) Bluetooth transceiver module. This device is a Bluetooth specification 2.0-support module that has an approximate range of 10-meters. The ECG acquisition hardware Bluetooth module is configured as a Master, and the mobile phone is considered to be functioning as a Slave. Figure 2 shows the ECG signals flow-chart.

Figure 1. Real-time ECG system block diagram

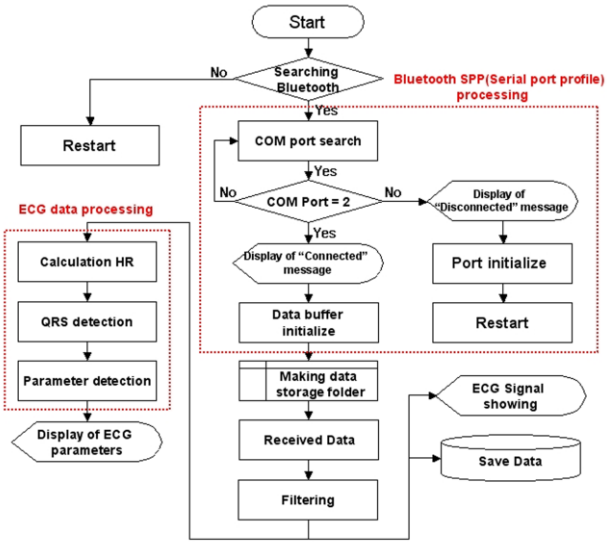


Figure 2. ECG signals receive flow chart

C. Data Management

A display of real-time ECG signals attempted on the screen display and data were saved in memory at the same time. The data were saved in binary format ASCII-code type, and rule of ECG data file name created follow; include Day, Time, a minute, and a second. For example, "xxxxxxxx.ecg" is an example of the filename of the data files saved. Also, the first time measurement of the data generated were saved in new storage folder in base driver at SmartPhone. This ECG data is available for administration through the history and management screen in the application program. It also protects patient information.

D. ECG Signal Processing

Recorded physiological signals usually have an original signal contaminated with noise. The noise is encountered at every stage of data acquisition until the data is digitized. Therefore, power noise, muscular contract noise, electrode movement with signal wandering, and analog-to-digital converter noise all perturb the ECG signals. If an electrode is removed the ECG signal becomes indecipherable. Power line interference noise is electromagnetic field from the power line, which causes 50 or 60 Hz sinusoidal interference. This noise causes problem in interpreting low amplitude waveform like ECG. Hence, many methods have been utilized on the removal of the power line interference in the ECG signals [7]. The wavelet coefficient threshold based hyper shrinkage function to remove power line frequency was used in [7], a nonlinear adaptive method to remove noise was used in [8], and subtraction procedure for power line interference removing from ECG which is extended to almost all possible cases of sampling rate and interference frequency variation was used in [9]. Power line noise cancellation based on these methods take a lot of operation time, as well as difficult to apply for a real time system. Therefore, we used an Infinite Impulse Response (IIR) notch filter. Though it has short processing time, it does not consider tracking frequency or removing a specific bandwidth rather than 60 Hz peak. This system is indented for real-time processing. The difference equation for this filter is as follows.

$$y[n]=\sum_{k=0}^{M} b_k\, x[n-k]+\sum_{k=1}^{N} a_k\, y[n-k] \tag{1}$$

Another type of unwanted signal in ECG is the baseline wander. Baseline wander can be caused by respiration, electrode impedance change and body movements. Baseline wander makes manual and automatic analysis of ECG recordings difficult, especially the measuring of ST-segment deviation, which is used for diagnosis of ischemia. Baseline wander elimination has been addressed in many different ways. The most widely used method uses cubic spline filtering and linear phase filtering for estimating the baseline drift [10][11]. We used baseline wander interference cancellation method based on band-pass sixth order Butterworth digital filter. The transfer function equation of the digital filter is shown below.

$$H(S)=\frac{B(S)}{A(S)}=\frac{b(1)S^{n}+b(2)S^{n-1}+.....+b(n+1)}{S^{n}+a(2)S^{n-1}+.....+a(n+1)} \tag{2}$$

E. QRS detection and Heart Rate Caculation

A typical ECG signal of a normal heartbeat can be divided into 3 parts, as depicted in Figure 3 [12], P wave or P complex, which indicates the start and end of the atrial depolarization of the heart; the QRS complex, which corresponds to the ventricular depolarization; and, finally, T wave or T complex, which indicates the ventricular depolarization [13][14]. QRS complex can be identified using general ECG parameter detection method. R-peak is easier to distinguish from noisy components since it has large amplitude. Noise and spike signals appear irregularly in ECG signals.

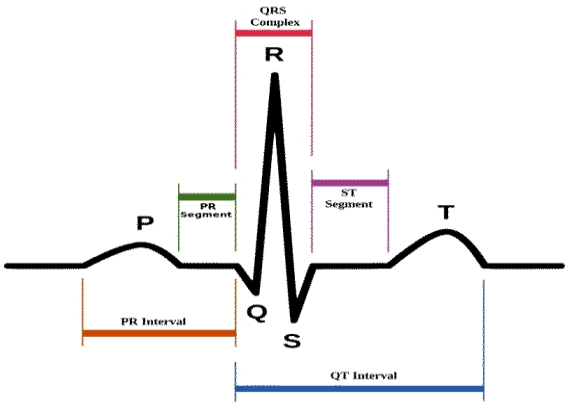


Figure 3. The general ECG waveform

After the pre-processing method, variable threshold method was used to further detect the R-peak. The formula for variable threshold value is defined as follows.

$$V_{TH} = [\, x(n) - x(n-1)\,] * 70\% \quad (3)$$

The threshold makes it possible to differentiate R peak from the baseline, which is corresponding to 70% of ECG peak data detection. We were able to find QRS complex based on the detected R-peak. Detection of QRS complex is particularly important in ECG signal processing. In our system, we used a robust real-time QRS detection algorithm [15]. This algorithm reliably detects QRS complexes using slope, amplitude, and other information. The information obtained from QRS detection, temporal information of each beat and QRS morphology information can be further used for the other ECG parameter detection. In order to detect QRS complex, the signal is initially passed through a band-pass filter. It is composed of cascaded high-pass and low-pass filters. Subsequent processes are five-point derivative (Eq. 4), square (Eq. 5), moving window integrator (Eq.6), and detection.

$$y(nT) = \frac{2x(nT) + x(nT-T) - x(nT-3T) - 2x(nT-4T)}{8} \quad (4)$$

$$y(nT) = [x(nT)]^2 \quad (5)$$

$$y(nT) = \frac{1}{N}[x(nT-(N-1)T) + x(nT-(N-2)T) + ... + x(nT)] \quad (6)$$

We computed instantaneous heart rate directly from R-R interval. In clinical settings, heart rate is measured in beats per minute (bpm). So the formula for determining heart rate from RR interval is given below (Eq. 7).

$$Heart\ Rate\ (bpm) = \frac{60{,}000}{RR\ Interval\ (ms)} \quad (7)$$

III. Experiments And Results

A. ECG Pre-processing

Appropriate shielding and safety consideration can be employed to reduce power line noise in addition to analog filtering as discussed in previous sections. After receiving signals at the receiver sides, it is preferred to remove this type of noise in the pre-processing step. Typically, band-stop (notch) filtering with cutoff, Fc = 50 or 60 Hz would suppress such a noise. Figure 4 illustrates the magnitude and phase response of a digital second order Infinite Impulse Response (IIR) notch filter with cutoff frequency of 60 Hz.

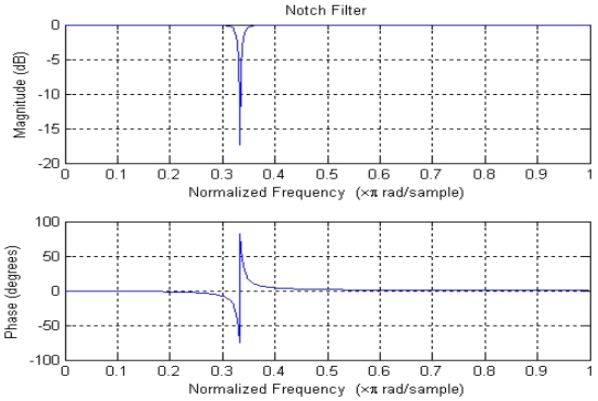


Figure 4. Magnitude and phase response of an Infinite Impulse Response(IIR) notch filter to remove 60Hz power line noise

The frequency of the baseline wander is usually below 0.5 Hz. This information particularly helps in the design of a high-pass filter in order to get rid of baseline wander. The design of a linear time-invariant high pass filter requires several considerations, most importantly, the choice of cut-off frequency and filter order. It is important to note that the ECG characteristic wave frequencies are higher than baseline wander. Therefore, carefully designed high pass filters with cut-off frequency 0.5 Hz can effectively remove the baseline. Baseline wander removing was performed using a band-pass sixth order Butterworth digital filter with cutoff 0.7-40Hz. To avoid distortion, zero phase digital filtering was performed by processing the data in both forward and reverse direction. In Figure 5, baseline wander correction using a linear digital filter is shown.

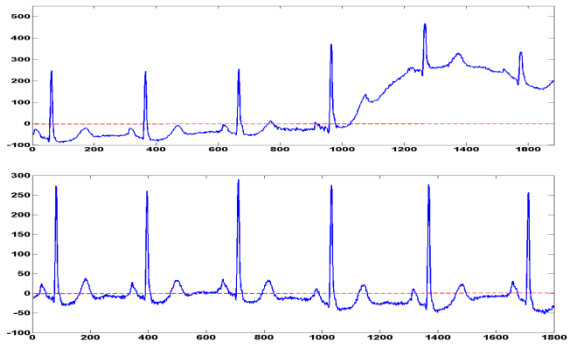


Figure 5. Baseline wandering correction (X-axis: samples, Y-axis: amplitude)

B. QRS detection using MIT-BIH ECG databases

The QRS detection provides the fundamentals for almost all automated ECG analysis algorithm. We tested the performance of the QRS detection on the MIT-BIH database, which is composed of half-hour recording of ECG of 48 ambulatory patients. The ECG recording "103.dat" shown in Figure 6 has been used to validate the algorithm and the following things are observed. The QRS detection algorithm [15] consists of several steps. First, the signal is passed through a digital band-pass filter. The pass band that maximizes the QRS energy is approximately in the 5-15 HZ range (Figure 6 (b)). Secondly, differentiation step is a standard technique for finding the high slop that normally distinguish the QRS complexes from other ECG waves (Figure 6 (c)); The squaring process makes the result positive and emphasize large differences resulting from QRS complexes (Figure 6 (d)); The moving window integration provides the slop and width of the QRS complex (Figure 6 (e)). The choice of window sample size is an important parameter. We choose window of 83 ms (i.e., 30 samples for a sampling frequency of 360 samples/s). Finally, an adaptive threshold was applied to identify the location of QRS complexes (Figure 6 (f), (g)).

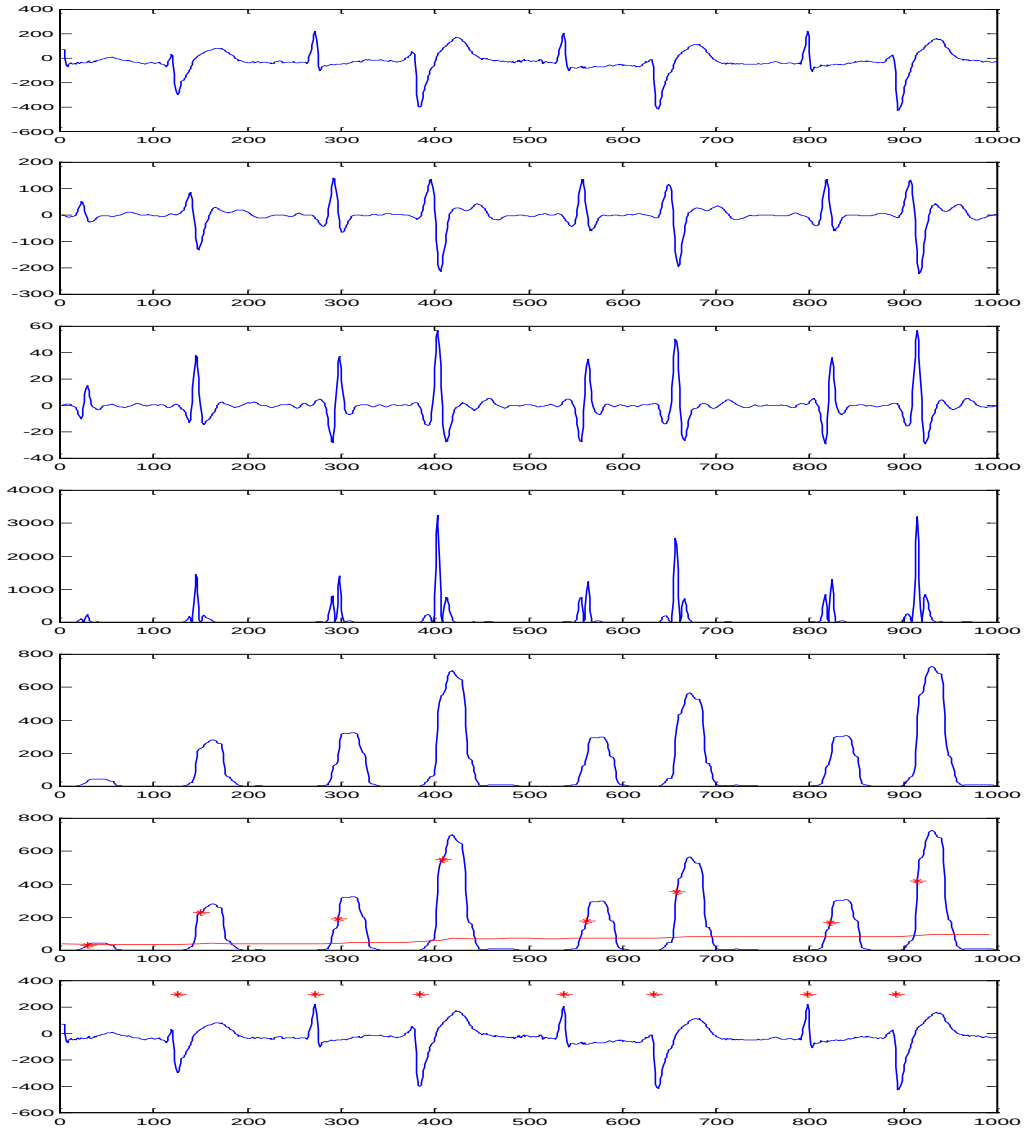


Figure 6. An example of QRS detection: (a) original ECG signal; (b) bandpass filter; (c) derivative; (d) square; (e) Moving window integrator; (f) QRS complex detection (X-axis: samples, Y-axis: amplitude)

C. Wireless Communication Test

Wireless communication module consists of master and slave. The master part transmitted after acquisition of ECG signal and amplification, and slave part received the signal and saved in data buffer. The master transmitted after conversion to digital value for 0 to 255 of the analog signal. Also, ECG hardware system and mobile phone communicate using Bluetooth. When the master/slave wireless module is first connected, the master module looks for a wireless module and attempts pairing for 5 to 10 seconds. If the ECG signal acquisition device and mobile phone are paired properly, the master module provides information to mobile phone wireless module with master address (Master Bluetooth module local address: 001901216B70) and starts the data transmission. The

device pairing success rate was around 85%, and the system requires initialization time for data buffer and screen display for 3 to 5 seconds. Communication test results between devices are shown in Table 1. We tested 20 cases. The master and slave module of average Pairing Time (PT) are 783 ms; ECG Signal Transmission Time (ECG STT) is 355.75 ms.

TABLE I. RESULT OF WIRELESS COMMUNICATION

Index	PT	STT	Remark	Index	PT	STT	Remark
1	1.320	None	Disconnect	11	597	412	Connect
2	865	508	Connect	12	1,160	None	Disconnect
3	867	489	Connect	13	883	385	Connect
4	843	416	Connect	14	914	353	Connect
5	1,241	None	Disconnect	15	758	402	Connect
6	536	452	Connect	16	670	352	Connect
7	632	455	Connect	17	822	379	Connect
8	545	394	Connect	18	598	514	Connect
9	582	406	Connect	19	615	386	Connect
10	601	386	Connect	20	608	426	Connect

a. PT: Pairing Time, STT: ECG Signal Transmission Time *(msec)*

D. Mobile Application Program

The mobile application program consists of five screen activities. These activities are main view, Bluetooth search view, real-time ECG signal view, signal parameter view, and data management view. Figure 7 shows the application program. The main view is the operation four-function key as shown in Figure 7 (a), each function key is moving to another functional screen and emergency connection button. The Bluetooth view searching to the ECG acquisition device on the mobile phone is shown in Figure 7 (b). Without this functionality the user has to stop the application program and restart the application. After the system connects successfully to the hardware, the visualization of the ECG signal and calculated heart rate graph is shown in Figure 7 (c). In case of not pairing, the program shows "Turn on ECG device" message box on the ECG signal view. A figure 7 (d) and figure 7 (e) is a user optional screen. Signal parameter view display calculated signal parameters: Heart Rate (HR), QRS duration, QT/QTc, PR and RR-interval. Data management view is displayed in case the measurement of ECG data has preserved for information. In order to search the previous ECG data the user has two options, either using drop list or calendar.

IV. CONCLUSIONS

The advances in mobile communication open up opportunities for developing mobile healthcare systems that monitor biomedical signals from patients. We developed such an ECG monitoring device for the advanced personal healthcare system using a mobile phone. The preliminary results showed a successful test of this mobile healthcare application. However, there are scopes of improvement, such as noise reduction, external memory expansion, memory space utilization, inclusion of more diagnostic parameters, and measurement of the physiological signals. Also, we would study the system safety for clinical trials in a variety of conditions. Above all, heart rate is a vital sign to determine the patient's condition and well-being. The heart rate monitoring tool should avoid any wrong results.

Finally, portable mobile healthcare has the potential to reduce long-term costs and improve quality of medical service, but it also faces many technical challenges. In future, more research on the small wireless electrical sensors and data compress technology for healthcare system is needed. The development of mobile health monitoring system would allow basic medical assessment of patients provided by medical staffs.

Figure 7. Mobile Phone Configuration: (a) monitoring system main view; (b) wireless connectionl; (c) real-time display of the ECG measurement; (d) the ECG parameters calculation value; (e) data history and management

ACKNOWLEDGMENT

Financial supports from North Dakota EPSCoR Grant #UND0014095 and University of North Dakota Faculty Research Seed Money #21418-4010-01843 are gratefully acknowledged.

REFERENCES

[1] B. A. Walker, A. H. Khandoker, & J. Black, "Low cost ECG monitor for developing countries", 2009 Fifth International conference on Intelligent Sensors, Sensor Networks and Information processing (ISSNIP), pp 195-200, December 2009.

[2] H, E. Sheref, S, Pham, N. E. Sherif, and E. Care, "Clinical evaluation of ECG data compression techniques for ambulatory recording", IEEE Conference on Engineering in Medicine and Biology, pp.1306-1307, November 1994.

[3] E. Jovanov, T. Martin, and D. Raskovic, "Issues in wearable computing for medical monitoring application: a case study of a wearable ECG monitoring device," The Forth International Symposium. Wearable Computers(ISWC), pp.43-49, October 2000.

[4] K. Y. Kong, C. Y. Ng, and K. Ong, "Web-Based Monitoring of Real-Time ECG Data," Computers in Cardiology 2000, pp. 189-192, Semptember 2000.

[5] J. R. Chang Chien, and C.C. Tai, "A new wireless type physiological signal measuring system using a PDA and the bluetooth technology," Biomedical Engineering: Applications, Basis and Communication, Vol. 15, No. 5, pp.229-235, October 2005.

[6] D. Kammer, G. McNutt, and B. Senese, "Bluetooth Application Developer's Guide", Syngress Publishing, Rockland, Mass, USA, 2002.

[7] S. Pooranchandra and N. Kumaravel, "A novel method for elimination of power line frequency in ECG signal using hyper shrinkage function," Digital Signal Processing, Vol. 18, No. 2, pp.116-126, March 2008.

[8] A. K. Ziarani and A. Konard, "A nonlinear adaptive method of elimination of power line interference in ECG signals," IEEE

Transaction Biomedical Engineering, Vol. 49, No. 6, pp.540-547, June 2002.

[9] G. Mihov, I. Dotsinsky, and T. Georgieva, "Subtraction procedure for powerline interference removing from ECG: improvement for non-multiple sampling," Journal of Medical Engineering & Technology, Vol. 29, No. 5, pp.238-243, September-October 2005.

[10] C. R. Meyer and H. N. Keiser, "Electrocardiogram baseline noise estimation and removal using cubic splines and state space computation techniques," Computers and Biomedical Research, Vol. 10, No. 5, pp.459-470, October 1997.

[11] J. A. Van Alste, W. Van Eck, and O. E. Herrmann, "ECG baseline wander reduction using linear phase filters," Computers and Biomedical Research, Vol. 19, No. 5, pp.417-427, 1986.

[12] Wikipedia, "Schematic diagram of normal sinus rhythm for a human heart as seen on ECG," January 2007,

[13] Available: http://en.wikipedia.org/wiki/File:SinusRhythmLabels.svg

[14] A. D. Jurik, J. F. Bolus, A. C. Weaver, B. H. Calhoun, and T. N. Blalock, "Mobile health monitoring through biotelemetry," The Fourth International Conference on Body Area Networks, April 2009.

[15] D. P. Coutinho, A. L. N. Fred, and M. A. T. Figueiredo, "One-lead ECG based personal identification using Ziv-Merhav cross parsing," 20th International Conference on Pattern Recognition, pp.3858-3861, August 2010.

[16] J. Pan, & W. J. Tompkins, "A real-time QRS detection algorithm," IEEE Transactions on Biomedical Engineering, Vol. 32, No. 3, pp.230-236, March 1985.

[17] L. Biel, O. Pettersson, L. Philipson, and P. Wide, "ECG analysis-a new approach in human identification," IEEE Transactions on Instrumentation and Measurement, Vol. 50, No. 3, pp.808-812, June 2001.

[18] Massachusetts Institute of Technology. MIT-BIH ECG database. Available: http://ecg.mit.edu/.

[19] W. Holsinger, K. Kempner, & M. Miller, "A QRS preprocessor based on digital differentiation", IEEE Transaction on Biomedical Englneering, Vol. 18, No, 3, pp.212-217, May 1971.

[20] S. Mallat, & W. Hwang, "Singularity detection and processing with wavelets", IEEE Transactions on information theory, Vol. 38, No. 2, pp.617-643, March 1992.

[21] S. Kadambe, R. Murray, & G. B. Bartels, "Wavelet transform-based QRS complex detector", IEEE Transactions on Biomedical Engineering, Vol. 46, No. 7, pp.838-848, July 1999.

[22] M. Bahoura, M. Hassani, & M. Hubin, "DSP implementation of wavelet transform for real time ECG wave forms detection and heart rate analysis", Computer methods and programs in biomedicine, Vol. 52, No. 1, pp.35-44, January 1997.

[23] J. Sahambi, S. Tandon, & R. Bhatt, "Using wavelet transforms for ECG characterization An on-linedigital signal processing system", IEEE Engineering in Medicine and Biology Magazine, Vol. 16, No. 1, pp.77-83, January-February 1997.

AUTHORS PROFILE

Duck Hee Lee received the MS degree in biomedical engineering from the Hanyang University, Seoul, South Korea in 2004. From 2005 to 2009, he worked at the Biomedical Engineering Division of the National Cancer Center (NCC), South Korea, developing a surgical robot system. In 2010, he was appointed Researcher of the University of North Dakota, Biomedical Signal Processing Laboratory, USA. Since 2004, he has worked on biomedical engineering research fields. His research interests include medical device and instrument, biomedical signal processing, and surgical robotics. He authored and co-authored more ten articles journals, conference proceedings and book chapter. He is a member of Korea Society for Medical and Biological.

Ahmed Rabbi received his B.Sc. and M.S. in Applied Physics, Electronics and Communication Engineering from the University of Dhaka, Bangladesh in 2004 and 2006 respectively. From 2007 to 2009 he worked as a telecom switiching engineer at Alcatel-Lucent Bangladesh coming to UND for graduate study. Currently, he is a Ph.D. student at the department of Electrical Engineering, University of North Dakota, USA. His research interests are biomedical signal and image processing, EEG signal processing, EEG-movement artifacts detection and filtering, epileptic seizure detection and prediction, and human performance assessment using EEG signals. He has participated as program committee member of an international conference. He has published over ten articles in refereed journals, conference proceedings and co-authored a book chapter. He is an active member of the IEEE and IEEE Engineering in Medicine and Biology Society (EMBS).

Jaesoon Choi received the Ph.D. degree in biomedical engineering from the Seoul National University, Seoul, South Korea in 2003. Since 2003, he has worked on biomedical engineering research fields. In 2011, he was appointed Research Professor of the Korea Artificial Organ Center (KAOC), Seoul, South Korea. He was responsible for various national and international research projects focused on key components for surgery robot system. His research interests include medical device and instrument, medical fusion multi-modal simulation, Vision-Haptic-Integrated Control Mechanism, and surgical robotics. He authored and coauthored more than 30 articles and holds ten patents. He is a member of Korea Society for Medical and Biological, Institute of Electrical and Electronics Engineers (IEEE), and International Society for Pediatric Mechanical Cardiopulmonary Support.

Reza Fazel-Rezai received his BSc. and M.Sc. in Electrical Engineering and Biomedical Engineering in 1990 and 1993, respectively. He received his Ph.D. in Electrical Engineering from the University of Manitoba in Winnipeg, Canada in 1999. From 2000 to 2002, he worked in industry as a senior research scientist and research team manager. Then, he joined academia at Sharif University of Technology and later the University of Manitoba as Assistant Professor in 2002 and 2004, respectively. Currently, he is Assistant Professor and the Director of Biomedical Signal Processing Laboratory at the Department of Electrical Engineering, University of North Dakota, USA. His research interests include biomedical engineering, signal and image processing, brain computer interface, EEG signal processing, seizure detection and prediction, neuro-feedback, and human performance evaluation based on EEG signals.

Development of knowledge Base Expert System for Natural treatment of Diabetes disease

Sanjeev Kumar Jha
University Department of Mathematics,
Babasaheb Bhimrao
Ambedkar Bihar University, Muzaffarpur - 842001
Bihar (India)

D.K.Singh
University Department of Mathematics,
Babasaheb Bhimrao
Ambedkar Bihar University, Muzaffarpur - 842001
Bihar (India)

***Abstract*—The development of expert system for treatment of Diabetes disease by using natural methods is new information technology derived from Artificial Intelligent research using ESTA (Expert System Text Animation) System. The proposed expert system contains knowledge about various methods of natural treatment methods (Massage, Herbal/Proper Nutrition, Acupuncture, Gems) for Diabetes diseases of Human Beings.**
The system is developed in the ESTA (Expert System shell for Text Animation) which is Visual Prolog 7.3 Application. The knowledge for the said system will be acquired from domain experts, texts and other related sources.

Keywords- Expert System; ESTA; Natural treatment; Diabetes.

I. INTRODUCTION

This article presents the conceptual framework of natural treatment methods available for diabetes. The main goal of this research is to integrate all the natural treatment information of diabetes in one place.

Expert System named as Sanjeevani is developed using ESTA (Expert System Shell for Text Animation) as knowledge based system to describe the various Natural therapy methods for treatment of Diabetes disease and various other diseases.

The main purpose of the present study is in the design and development of an expert system which provides the information of different types of natural treatment (Massage, Acupuncture, Herbal/Proper Nutrition and gems) of Diabetes. The system background starts with the collection of information of different methods of treatment available for Diabetes diseases. The acquired knowledge is represented to develop expert System.

II. DEVELOPMENT OF EXPERT SYSTEM

We are in the process of development of Sanjeevani Natural therapy expert system that is developed in ESTA Application. It is designed for assisting in treatment of the Diabetes disease and various other diseases which can be cure naturally with different methods (Massage, Herbal/Proper Nutrition, Acupuncture, Gems) and provide treatment solutions.

The natural therapy consists of a variety of natural body therapies, soul therapies and energy therapies for healing, that we found helpful for peoples health and wellness and which don't cost the earth

The natural treatment is the process of healing/curing diseases through natural, drugless and the most harmless process. Human beings are an intrinsic part of the nature. Therefore, nothing except nature can facilitate a complete cure for human machinery disorders.

These are the various methods of Natural therapy:--

- Massage
- Acupuncture
- Herbal/Proper Nutrition
- Gems

Diabetes Diseases
Treatment Methods
1) Natural Care(Herbal / Proper Nutrition) 2) Acupuncture 3) Homeopathic 4) Massage 5) Gems
Treatment Solutions Advice

Figure 1: Representation of Natural Treatment of Diabetes Diseases

III. DESIGN OF EXPERT SYSTEM

The Expert System can be created by using ESTA by building the knowledge base.

Expert System --→ ESTA + Knowledge Base

The Quality of Expert System is depends on its knowledge base.

The process of developing knowledge base is:-

a) Identifying the input of Problem
b) Gaining Knowledge
c) Representation of Knowledge

A. Identifying the Input of Problem

For developing the expert system first we have to identify the problem and its behaviors.

The Input for our system is regarding identifying the different types of natural treatments (Massage, Acupuncture,

Herbal/Proper Nutrition and gems) available for Diabetes Disease.

B. Gaining Knowledge

Gaining Knowledge is very important in developing the expert system. The acquired gained knowledge is analyzed and processed to give the best solution of the problem.

The knowledge of different types of Natural treatments (Massage, Acupuncture, Herbal/Proper Nutrition and gems) of Diabetes has been gained by reading books, browsing Internet and also got information from consultation of Physician in their respective areas.

C. Representation of Knowledge

Representation of knowledge is the last phase of the development of knowledge base system. There are various approaches for representation of Knowledge into knowledge base.

Each Knowledge base contains rules for a specific domain. Thus, for a natural treatment of diabetes expert system the knowledge base will contain rules relating certain natural treatment methods of diabetes such as Massage, Acupuncture, Herbal/Proper Nutrition and gems.

ESTA has all facilities to write the rules that will make up a knowledge base. Further, ESTA has an inference engine which can use the rules in the knowledge base to determine which advice is to be given to the expert system user or to initiate other actions

Representation in ESTA is the rule based in logical paradigm of simple if-then rules in backward or forward chaining. We have chosen here the backward chaining for knowledge representation with simple if-do pair in place of if-then rules. Here we have considered two major knowledge representations namely Sections and Parameters. The top level of representation of knowledge in ESTA is section. It contains the logical rules that direct the expert system how to solve problem, actions to perform such as giving advice, going to other sections, calling to routines etc. The first section in ESTA is always named as start section. The advice is given when condition(s) in the section is (are) fulfilled. Parameters are used as variable and it determines the flow of control among the sections in the Knowledge Base. A parameter can be one of the four types: Boolean or logical, Text, Number and Category parameters.

We have developed the various Parameters and Sections for developing this expert system.

IV. Representation Of Sanjeevani Expert System

The Knowledge representation in ESTA is based on the items: a) Section b) Parameters c) Title

A. Representation of Parameters Used in Developing Expert System

In FIGURE 1: Here we have defined the disease parameter which is of type category describing the various types of disease.

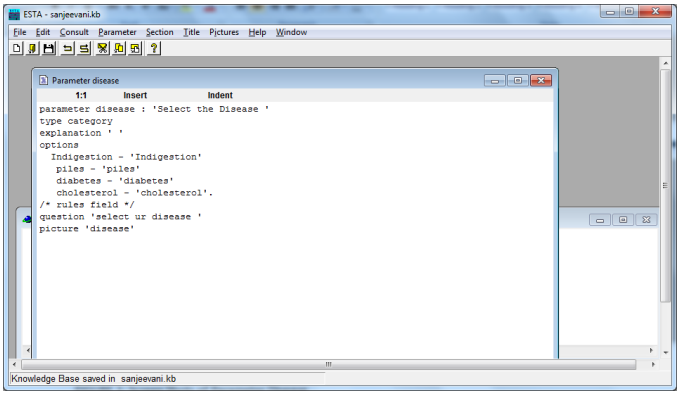


Figure 1. Screen Shots of Parameter Disease

In FIGURE 2: Here we have defined the diabetesop parameter which is of type category describing the various types of natural treatment available for diabetes disease.

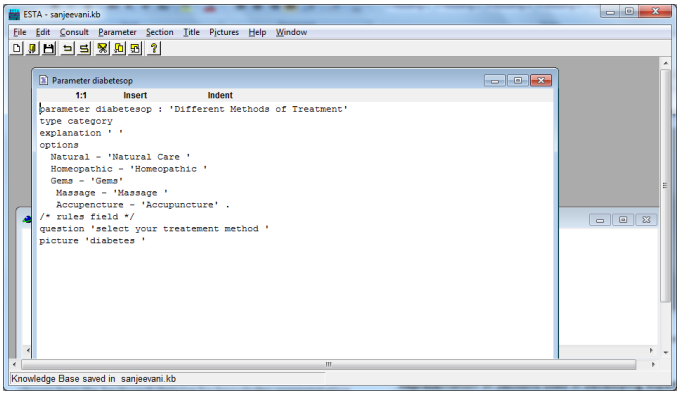


Figure 2. Screen Prints of Parameter diabetesop

B. Representation of Sections used in developing expert system

In FIGURE 3: Here we have a main section start which is developed to transferring controls in accordance with the user's response about disease

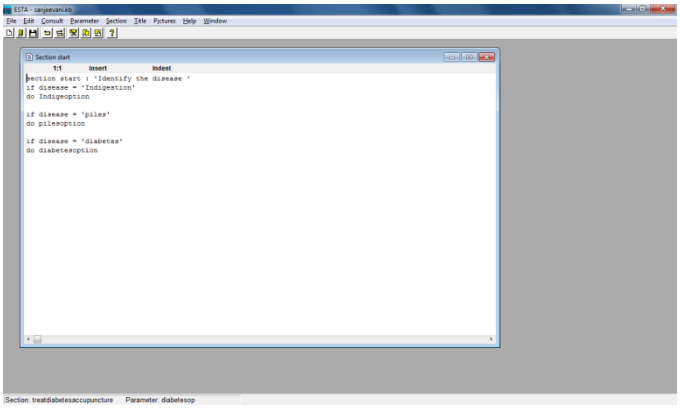


Figure 3. Screen Print of Start Section

In FIGURE 4: Here causeofdiabetes Section describes the diabetes disease and its symptoms

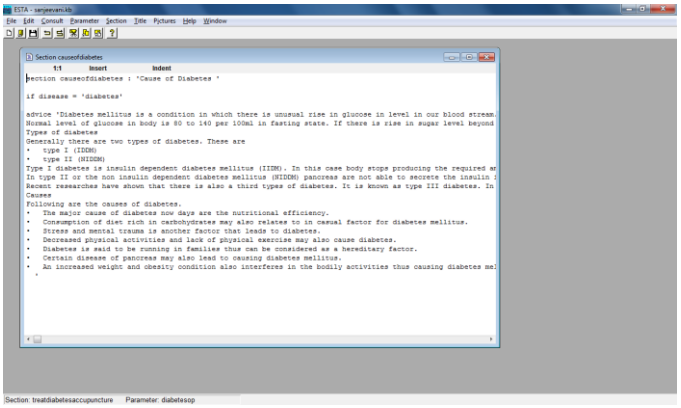


Figure 4. Screen Print of Causeofdiabetes Section

In FIGURE 5: Here in the diabetesoption section describes the various natural treatment options available for diabetes disease.

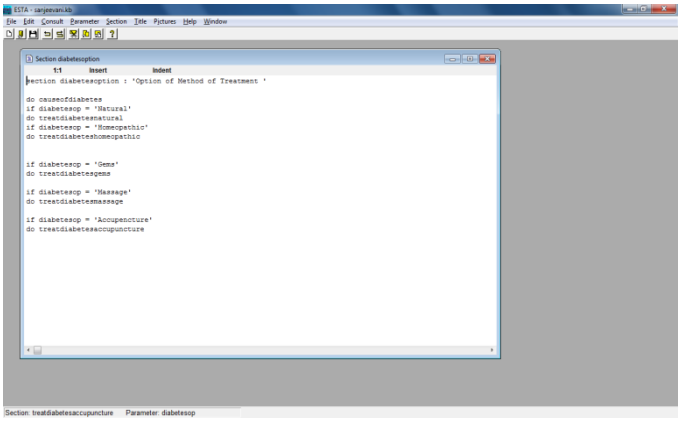


Figure 5. Screen Print of Section of diabetesoption

In FIGURE 6: Here in the treatdiabetesnatural section describes the Natural Care (Herbal / Proper Nutrition) treatment solution of diabetes disease

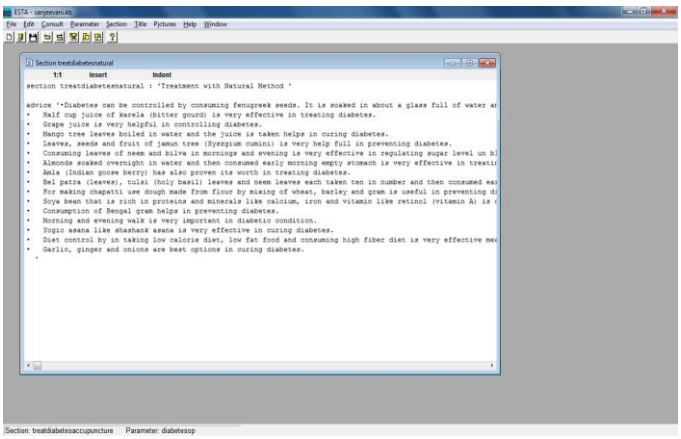


Figure 6. Screen Print of Section treatdiabetesnatural

1.1. Representation of Title

In FIGURE 7: It describes the Title of Sanjeevani expert system.

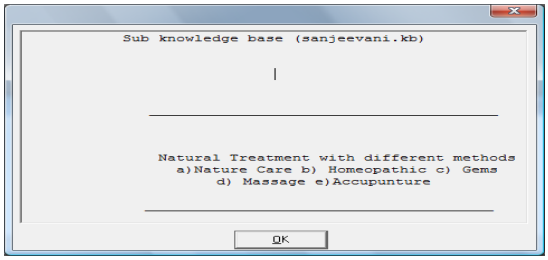


Figure 7. Screen Print of Title of expert System

V. Consultation Of Expert System

In FIGURE 8: It describes the beginning of Consultation of Sanjeevani Expert System. It will ask the users to select the disease (Diabetes) for which they want different type of natural treatment solution.

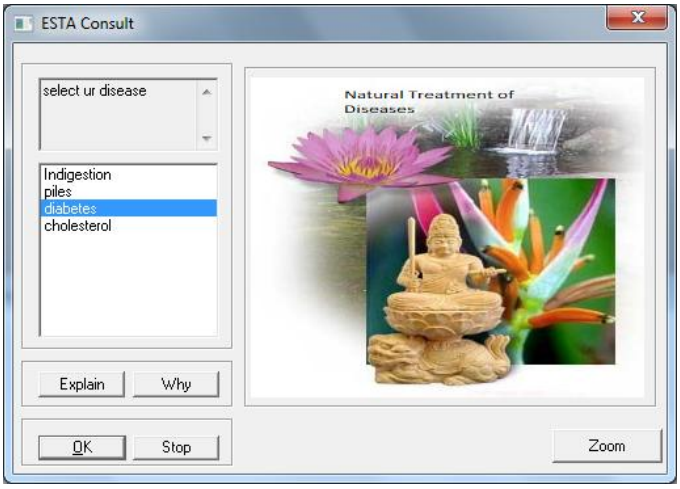


Figure 8. Screen Print of ESTA Consult of Diabetes

In FIGURE 9: It describes the diabetes diseases and its symptoms

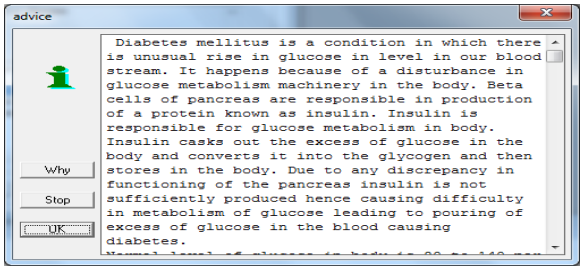


Figure 9. Screen Print of Diabetes description and Symptoms

In FIGURE 10: It describes the different types of Natural treatment methods available for Diabetes disease. It will ask users to select one of the Natural treatment methods for getting details treatment advice.

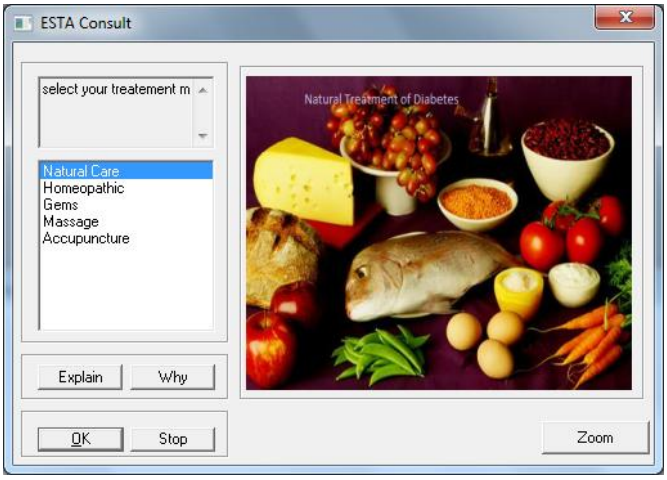


Figure 10. Screen Print of selection of Natural Treatment Method

In FIGURE 11: It describes the Natural Care (Herbal / Proper Nutrition) treatment solution of diabetes disease

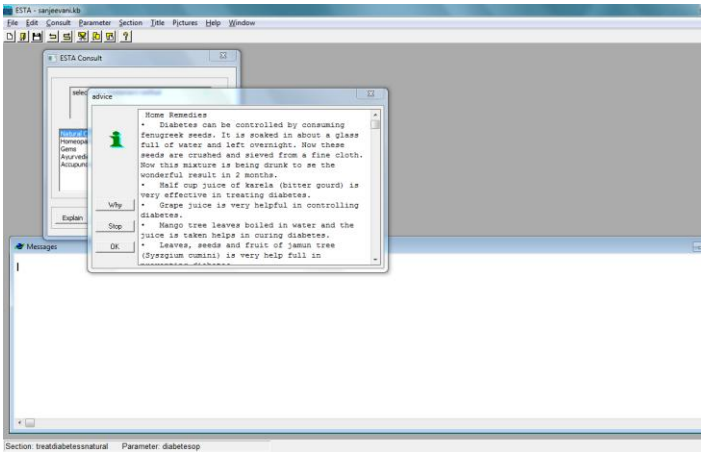


Figure 11. Screen Print of Natural Treatment Solution

VI. CONCLUSIONS

As the Sanjeevani project shows, applied work in developing of expert system describing the various Natural therapy methods for treatment of Diabetes disease and various other diseases.

The field of medical artificial intelligence is particularly appealing to the physicians and computer scientists working in the area.

The long-term challenges are well recognized-such as the need for mechanisms that will assure completeness and shared knowledge bases for different natural methods treatment of different diseases.

This Expert System development is one of the steps to get the integrated knowledge base system which will help in getting information of various natural treatment methods available for disease to general users (Patient and Physician)

REFERENCES

[1] Rajkishore Prasad, Kumar Rajeev Ranjan, and A.K. Sinha, "AMRAPALIKA: An expert system for the diagnosis of pests, diseases, disorders in Indian mango," Knowl.-Based Syst. 19(1): 9-21 (2006).

[2] Francisco Elazegui, and Zahirul Islam, "Diagnosis of common diseases of rice," © 2003, International Rice Research Institute.

[3] S.S.Patil, B.V.Dhandra, U.B.Angadi, A.G.Shankar, and Neena Joshi, "Web based Expert System for Diagnosis of Micro Nutrients Deficiencies in Crops", Proceedings of the World Congress on Engineering and Computer Science 2009 Vol I WCECS 2009, October 20-22, 2009, San Francisco, USA.

[4] Turban, E., Expert Systems and Applied Artificial Intelligence. New York: Macmillan Publishing Company, 1992

[5] Prolog Development Center, Expert System Shell for Text Animation (ESTA), version 4.5, A/S. H. J. Volst Vej 5A, DK-2605 Broendby Denmark, copy right©1992-1998

Book

[6] Expert System: Principles and Programming by Joseph C.Giarratano

[7] PROLOG –Programming for Artificial Intelligence by Ivan Bratko

Software

[8] Visual Prolog 7.3 ESTA Application

AUTHOR'S PROFILE

Sanjeev Kumar Jha has received M.C.A from Madras University in 2002; Presently He is pursuing PhD from Ambedkar Bihar University, Muzaffarpur. He has a Software Industry experience of more than 7 years. His areas of interest include Artificial Intelligence, Expert System and Software Testing.

Maximum-Bandwidth Node-Disjoint Paths

Mostafa H. Dahshan
Dept. of Computer Engineering
College of Computer and Information Sciences
King Saud University
Riyadh, Saudi Arabia

***Abstract*—This paper presents a new method for finding the node-disjoint paths with maximum combined bandwidth in communication networks. This problem is an NP-complete problem which can be optimally solved in exponential time using integer linear programming (ILP). The presented method uses a maximum-cost variant of Dijkstra algorithm and a virtual-node representation to obtain the maximum-bandwidth node-disjoint path. Through several simulations, we compare the performance of our method to a modern heuristic technique and to the ILP solution. We show that, in a polynomial execution time, our proposed method produces results that are almost identical to ILP in a significantly lower execution time.**

Keywords-Maximum Bandwidth; Disjoint Paths; Widest Pair; Linear Programming; ILP; NP-Complete; Dijkstra Algorithm; Multiple Constrained Path; MCP.

I. INTRODUCTION

Path optimization is a fundamental problem in data networks. Traditional path optimization aims to find the single lowest-delay path between a given source and destination nodes. The main application for such a problem is routing in IP networks. For other applications, variants of the problem are needed. QoS requirements may set one or more constraints to be satisfied along the path [1]. In general, QoS constraints can be classified into additive, multiplicative, and concave. Additive constraints include: delay, jitter, and hop count. Multiplicative constraints include the probability of packet arrival and link reliability. Concave constraints include finding the minimum or maximum bandwidth along the path to represent the bandwidth of the path [2]. In addition to single path QoS requirements, recovery plans may require having one or more backup paths to be ready in case the primary path fails. Multiple paths are also required in traffic engineering schemes to provide load balancing [3], [4]. Multiple paths usually have an additional constraint to be link-disjoint or node-disjoint. Node-disjoint paths are usually harder to find but provide more robustness in case of node failures.

The complexity of path optimization varies depending on the type and number of constraints. Several studies have shown that the Multiple Constrained Path (MCP) problems are generally NP-complete and are not solvable in polynomial time [5], [6]. Furthermore, finding disjoint paths with a single constraint is generally an NP-complete or NP-hard problem [7], [8].

In this paper, our focus is on the problem of finding two node-disjoint paths such that the bandwidth sum of the two paths is the maximum possible two-disjoint-paths sum between a given source and destination nodes in the network. This is essentially an MCP problem with two constraints: The first constraint is for the two paths to be node-disjoint. The second constraint is maximizing the bandwidth sum of the two paths. This is also an NP-complete problem as shown in [9]. We develop a near-optimal method for solving this problem in polynomial time. The proposed method uses a virtual-node representation from the original network. We implement a variant of Dijkstra algorithm that finds the optimal path based on the maximum bandwidth [10]. The variant algorithm is further modified to work concurrently on two paths, avoiding nodes that lead to overlapped paths in the original network. The algorithm is then applied iteratively to obtain the maximum disjoint path in the actual network.

In the remaining part of the paper, we discuss related studies that attempted to find solutions to the maximum-pair disjoint paths and similar problems. Next, we illustrate the modified Dijkstra algorithm that finds the maximum-bandwidth path. The new method is then presented in details and demonstrated by an example. The performance of our method is evaluated and compared to a modern heuristic algorithm and to the exact solution using ILP. Analytical study of the presented method is then presented to show the order of its execution time. The paper is concluded with a summary and future work.

II. RELATED WORK

The problem of finding disjoint paths has been subject to intensive research. In particular, several studies have addressed the problem of finding maximum combined bandwidth in disjoint paths. Shen and Sen [9] have discussed two versions of the problem. The first version is finding pair of disjoint paths with maximum combined bandwidth which they call "widest pair of disjoint paths". The second version is finding a pair of disjoint paths such that the bandwidth of the first path is greater than or equal to X1 and the bandwidth of the second path is greater than or equal to X2. They proved that both versions of the problem are NP-complete and provided both exact solutions using ILP and two approximate heuristic solutions.

The first solution, deterministic heuristic algorithm (DHA), works in two phases: First, it uses a relaxed version of ILP, in which solutions are not necessarily integers. If it produces integer values, then the solution is accepted. If not, the second phase replaces the capacity of each edge with values from the solution obtained in the phase 1 and then applies Suurballe's algorithm [11], [12] to attempt finding the two paths. The second solution, randomized heuristic algorithm (RHA), uses

the same phase 1 as in DHA. The second phase constructs two graphs g1 and g2 with edges assigned values obtained from ILP in phase 1, first path for g1 and second path for g2. Then, it uses random walks to get the two disjoint paths by alternating the random function between g1 and g2. Despite the relaxation of ILP in these heuristics, the time required for ILP execution is generally in exponential order.

Shen et al [13] have addressed the problem of maximizing bandwidth through disjoint paths. They proposed a heuristic algorithm called Algorithm-1. The algorithm generates all possible paths from source to destination using the algorithm of Am et al [14]. The algorithm then creates a path intersection graph G' in which each node represents one of the paths obtained in step 1. Paths are considered intersected if they share at least a common edge. Nodes of the G' have weights to indicate the bandwidth of the path. Finally, the algorithm finds the maximum weight independent set S of G', which are nodes in G' with maximum weights and which are not connected by edges. Finding maximum independent set is an NP-complete problem and thus approximate algorithms are used.

Leng et al [15] have studied a different but related problem of finding shortest pair of disjoint paths with bandwidth guarantee (SPDP-BG). In this problem, it is required to find a pair of disjoint paths with minimum cost, while guaranteeing a minimum bandwidth of a defined value X. They proved that the SPDP-BG problem is NP-complete and presented a heuristic algorithm to solve the problem. Their heuristic algorithm finds a pair of disjoint paths with guaranteed minimum bandwidth, and then modifies them to gradually minimize their lengths. The algorithm first finds the widest-bandwidth path.

Next, it uses the aforementioned DHA algorithm of [9] to get the widest pair of disjoint paths. The length of the two disjoint paths is used as an upper bound for the path length. After that, find k-shortest paths using the algorithm of [16]. Next, loop in each of the k-shortest paths in ascending order. For each path, find a second disjoint path using Dijkstra algorithm such that the combined bandwidth of the two paths is at least X. If the length of the two found paths is less than the previous upper bound, set the current path length as an upper bound, and continue until all k-paths have been examined.

Loh et al [17] have addressed a more general version of the problem – finding multiple disjoint paths between source and destination nodes. They propose a polynomial-time heuristic algorithm, Maximum Bandwidth Algorithm (MBA), for solving this problem. The MBA algorithm creates two sets of edges. One set, called ES, contains edges ongoing from the source, and the other set, BS, contains all other edges. In both sets, the edges are sorted in descending order based on their bandwidth. At each round, take the highest-bandwidth edge in ES and remove all other edges in the network with lower bandwidths. Then, attempt to get the path with maximum bandwidth using Dijkstra algorithm. If no path is found, take the edges in BS in descending bandwidth, removing all edges from the network with lower-than-current bandwidth and run Dijkstra again. Before running Dijkstra algorithm, the cost of each edge is set to the result of subtracting the bandwidth of the link from a fixed number larger than the maximum bandwidth from the previous step. After finding the first maximum-bandwidth path, remove all the edges in that path from the network and continue to the next round with the next highest-bandwidth edge in ES. The authors have shown that the MBA algorithm produces the optimal disjoint paths in 99% of the cases using only 0.005% of the CPU time required using the optimal, but exponential, brute-force (BF) algorithm. Since it is one of the newest developed algorithms for the problem of concern, we have used the MBA algorithm for comparison with the algorithm proposed in this paper.

III. The Modified Dijkstra Algorithm

Sahni et al [10] have developed a modified version of Dijkstra algorithm to calculate the maximum bandwidth from a given source node s to a given destination node d. The algorithm shown in Figure 1 is based on Sahni's algorithm, but is extended to find the maximum bandwidth from a source node s to all other nodes in the network. The algorithm is also adapted to our syntax.

```
01 Algorithm MaxBandwidth(s)
02 for i = 1 to n do
03   if node[i] is a neighbor of s then
04     set maxbw of node[i] =
                bandwidth of link[s, i]
05   else
06     set maxbw of node[i] = 0
07   end if
08   set previous of node[i] = s
09 end for
10 set previous of node[s] = 0
11 label node[s] as PERMANENT
12
13 while exists node[i] with label = TENTATIVE
   do
14   find x = i with maxbw of node[i]
              not labeled as PERMANENT
15   if maxbw of node[x] = 0 then
16     exit #no more paths
17   else
18     label node[x] as PERMANENT
19   end if
20
21   for each neighbor node[v] of node[x] do
22     if  node[v] is not PERMANENT then
23       if minimum (maxbw of node[x],
            bandwidth of link [x,v])
            > maxbw of node[v]
24       then
25         set previous of node[v] = node[x]
26         set maxbw of node[v] =
                  minimum(maxbw node[x],
                  bandwidth of link[x, v])
27       end if
28     end if
29   end for
30 end while
```

Figure 1. The modified Dijkstra algorithm

The main differences between the modified algorithm and Dijkstra algorithm are explained. First, the search function in the modified algorithm searches for the maximum-bandwidth rather than the minimum-cost node (14-19). Second, the relaxation (neighbor update) phase chooses the largest between the bandwidth of the neighbor node and the minimum of the link bandwidth and the bandwidth of the current node (23-28). This is different from Dijkstra's relaxation phase, which chooses the minimum between the cost of the neighbor node and the sum of the link cost and the cost of the current node.

IV. The Proposed Algorithm

The algorithm presented in this paper can be summarized as follows. Work on finding two paths concurrently: one path is called R (red path) and the other is called B (blue path). At each instance of the algorithm execution, the algorithm finds the R path with maximum bandwidth, together with a node-disjoint path B with bandwidth not less than a specified limit. We will call the algorithm MLBDP, short for Max-Limit Bandwidth Disjoint Path.

Our algorithm uses the aforementioned modified Dijkstra algorithm to work on dual paths concurrently. To facilitate this approach, a virtual representation of the network is created with n×n virtual nodes (vnodes), where n is the number of nodes in the original network. Each virtual node is denoted by two symbols which represent the two current nodes in the dual path.

```
01  Algorithm MLBDP (s, limit)
02  #s is the source node in the original G
03  #limit is the minimum bandwidth allowed in
    the R path
04  #ss is source node in the virtual network
05  #n is the number of nodes in the original
    network G
06
07  for i = 1 to n do
08    for j = 1 to n do
09      visited list of vnode[ij] =
          new empty list
10      label vnode[ij] as TENTATIVE
11      set R.maxbw of vnode[ij] = 0
12      set B.maxbw of vnode[ij] = 0
13    end for
14  end for
15
16  for i = 1 to n do
17    label vnode[is] and vnode[si] as
        PERMANENT
18  end for
19
20  for i = 1 to n do
21    for j = 1 to n  do
22      if node[i] ≠ node[j] are neighbors of s
        and bandwidth of link[s,j] ≥ limit then
23        set R.maxbw of vnode[ij] =
                bandwidth of link [s,i]
24        set B.maxbw of vnode[ij] =
            bandwidth of link [s,j]
25        set previous of vnode[ij] =
            vnode[ss]
26      end if
27    end for
28  end for
29
30  #loop until reach all the destinations
     with two paths
31  while exists vnode[ii] with label =
      TENTATIVE do
32    #find vnode with maximum R.maxbw and
        make it PERMANENT
33    repeat
        find xy = ij with maximum R.maxbw of
34        vnode[ij] not labeled as PERMANENT
35      if R.maxbw of vnode[xy] = 0 then
36        exit #no more paths
37      else
38        label vnode[xy] as PERMANENT
39    until x ≠ y
40
41    for each neighbor node[v] of node[x] do
42      if  vnode[vy] is not PERMANENT then
43        if minimum (R.maxbw of vnode[xy],
            bandwidth of link [x,v]) >
            R.maxbw of vnode[vy]
44           and link[x, v] not in
               visited list of vnode[xy]
45        then
46          set previous of vnode[vy] =
              vnode[xy]
47          set visited list of vnode[vy] =
              visited list of vnode[xy]
              + link[v,x]
48          set R.maxbw of vnode[vy] =
              minimum(R.maxbw vnode[x,y],
              bandwidth of link[x,v])
49          set B.maxbw of vnode[vy] =
              B.maxbw of vnode[xy]
50        end if
51      end if
52    end for
53
54    for each neighbor node[u] of node[y] do
55      if  vnode[xu] is not PERMANENT  then
56         if R.maxbw of vnode[xy] ≥
             R.maxbw of vnode[xu]
57           and link[u, y] not in
                visited list of vnode[xy]
58           and minimum(B.maxbw of vnode[xy],
                bandwidth of link [y,u])
                  ≥ limit
59        then
60          set previous of vnode[x,u] =
              vnode[xy]
61          set visited list of vnode[xu] =
```

```
            visited list of vnode[xy]
            + link[u,y]
62        R.maxbw of vnode[xu] =
            R.maxbw vnode[x,y]
63        B.maxbw of vnode[xu] =
            minmum( B.maxbw of vnode[x,y],
            bandwidth of link[y,u])
64      end if
65    end if
66  end for
67 end while
```

Figure 2. The proposed MLBDP algorithm

The algorithm starts by initializing the R and B bandwidths of all virtual nodes to 0 and labeling them as TENTATIVE (7-14). The source virtual node [ss] which corresponds to the source, and any virtual node [si] or [is], are marked as PERMANENT (16-18).

Next, find all permutations of two nodes i and j which are neighbors of s such that the bandwidth of the link between the source and the second node j is at least equal to the limit. The maxbw of the R path (R.maxbw) is set to the bandwidth of link[s, i] and the maxbw of B path (B.maxbw) is set to the bandwidth of link[s,j] (20-28).

The main loop of the algorithm is next started (30-end), which remains until all virtual nodes [ii] are marked as PERMANENT. Note that reaching a virtual node with identical two indexes [ii] from the source means that the destination node with the corresponding single index [i] has been reached with two disjoint paths.

The repeat loop in (30-39) performs the search for R.maxbw vnode (vnode with max bandwidth R path) with maximum bandwidth greater than 0, if such vnode is found, it will be marked as PERMANENT. This is similar to the search phase in the modified Dijkstra algorithm. If the vnode [xy] has two identical constituents (x equals y), one destination is reached and so the algorithm doesn't examine the neighbor of that vnode, but continues to search for the next R.maxbw vnode. The loop will end when a vnode [xy] with $x \neq y$ is found. The vnode [xy] becomes the current working vnode.

The for loop in (41-52) iterates on all neighboring nodes [v] of the current working node [x] (first side of the path), as done in the relaxation phase of the modified Dijkstra algorithm. The minimum of the R.maxbw (current max bandwidth of R path) of the current node and the bandwidth of the link[x, v] is compared with the R.maxbw (current max bandwidth of R path) of virtual node [vy]. If this minimum is greater, this means a larger R.maxbw is found for the virtual node [vy]. Thus, the current R.maxbw of vnode [vy] is set to this minimum. The for loop in (54-66) iterates on all neighboring nodes [u] of the current working node [y] (second side of the path). It is similar to the previous for loop, except that the minimum of B.maxbw and the bandwidth of the link [y, u] is compared with the specified limit parameter, instead of the current max bandwidth of the B path.

In both loops it is important to ensure that the path is node-disjoint. This is accomplished by maintaining a visited list for each virtual node. This list contains nodes which have been traversed in the current pair of paths. The visited list is checked before a new node is added to the path (44, 57) and updated after the node is added to the path (47, 61).

Recall that the algorithm MLBDP finds the R path with maximum bandwidth, together with the B path with bandwidth ≥ limit. In order to find the node-disjoint path with the maximum total bandwidth, the algorithm MLBDP needs to be executed q times, where q is the number of unique link bandwidths. In each execution, the limit is set to one of the bandwidth values.

V. Example of the Proposed Algorithm

To demonstrate how the MLBDP algorithm works, consider the network shown in Figure 3. It is required to find the maximum-bandwidth node-disjoint path from a to d. We will explain a single run of MLBDP with limit = 7.

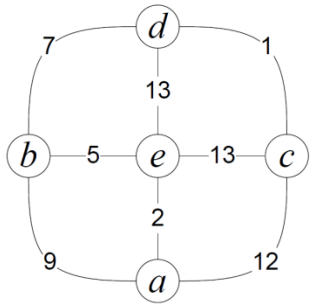


Figure 3. Example network to explain the MLBDP algorithm.

The algorithm will start by initializing the source virtual node aa as PERMANENT and initializing the (R.maxbw, B.maxbw) values of the virtual nodes adjacent to aa, as follows: eb = (2, 9), ec = (2, 12), bc = (9, 12) and cb = (12, 9). Note that vnodes ce and be will not be initialized and will remain at (0, 0) because the bandwidth of the Blue-side link (B.maxbw) is less than the limit of 7. The vnode with largest R.maxbw is cb (12, 9) with R.maxbw = 12, so it will be chosen as current vnode and labeled as PERMANENT. From cb, the neighbor vnodes are eb (12, 9), db (1, 9) and cd (12, 7). Again, ce (12, 5) will not be considered because B.maxbw of ce = 5 < limit. Both eb (12, 9) and cd (12, 7) have largest R.maxbw, but eb has a larger B.maxbw so the eb will be chosen as current vnode and labeled as PERMANENT.

From eb, the neighbor vnodes are bb (5, 9), db (12, 9) and ed (12, 7). ee will not be considered because its B.maxbw = 5. db (12, 9) will be chosen as current and labeled as PERMANENT. Neighbors of db are dd (12, 7) and bb (7, 9). Note that de will not be considered for two reasons. First, node e has been traversed in the path from aa to db (aa-cb-eb-db). Thus, the visited list of vnode db currently has the nodes a, c and e. The second reason is because B.maxbw of de = 5 < limit. The current vnodes with highest R.maxbw are cd (12, 7), ed (12, 7) and dd (12, 7). They will be marked successively. When vnode dd is marked, this means that node d has been reached with the node-disjoint path (aa-cb-eb-db). The constituent paths are (a-c-e-d), (a-b-d) and the combined bandwidth is 12+7 = 19, which is the maximum node-disjoint path from a to d.

VI. PERFORMANCE STUDY

As mentioned in the Introduction, the problem of finding maximum-bandwidth disjoint path is considered NP-complete. As such, the optimal solution of the problem can only be obtained using ILP. Execution time of ILP is exponentially proportional to the number of nodes in the network but is guaranteed to find all possible disjoint paths. Heuristic solutions, on the other hand, require much less time, but may not always find the maximum-bandwidth disjoint paths. The purpose of this section is to compare the performance of the MLBDP algorithm presented in this paper with both ILP and the MBA algorithm mentioned in Section II. The comparison is based on both the execution time and the maximum-bandwidth disjoint paths found.

A. ILP Formulation of the Problem

The ILP formulation of the problem of finding two disjoint paths has been developed in several related works. We use the formulation developed in [9]. Some modifications are made to the formulation to make it applicable for node-disjoint instead of link-disjoint paths.

Let V denote the set of nodes and E denote the set of links in the network. Links belonging to E are defined as pairs (u, v) that represent the nodes they are connecting. The source and destination nodes are denoted as s and t, respectively. The two paths are denoted as the red and blue paths. For each link (u, v) that belongs to E, four variables are defined: r(u, v), r(v, u), b(u, v) and b(v, u). These variables can take value 0 or 1. If the link from u to v belongs to the red path, r(u, v) = 1, else, r(u, v) = 0. Note that order of u of v is important, because it defines the direction of the path. Same can be said about b(u, v) and b(v, u). The bandwidths of the red and blue paths are denoted by yr and yb, respectively.

The function δ () is defined as follows:

$$\delta(x)=\begin{cases}1 & x=s\\ -1 & x=t\\ 0 & \text{otherwise}\end{cases}$$

With the previous definitions, the ILP formulation can be stated as follows:

Maximize $y_r + y_b$ such that:

$$\sum_{v\in V} r(x,v)-\sum_{u\in V} r(u,x)=\delta(x), \qquad \forall x\in V \quad (1)$$

$$\sum_{u\in V} b(x,u)-\sum_{u\in V} b(u,x)=\delta(x), \qquad \forall x\in V \quad (2)$$

$$\sum_{v\in V} r(v,x)+b(v,x)=2, \qquad \text{for } x=t \quad (3)$$

$$\sum_{v\in V} r(v,x)+b(v,x)=0, \qquad \text{for } x=s \quad (4)$$

$$\sum_{v\in V} r(v,x)+b(v,x)\le 1, \forall x\in V, \qquad x\notin\{s,t\} \quad (5)$$

$$y_r \le B_{uv}\cdot r(v,u)+M\cdot(1-r(v,u)), \quad \forall(u,v)\in E \quad (6)$$

$$y_b \le B_{uv}\cdot b(v,u)+M\cdot(1-b(v,u)), \quad \forall(u,v)\in E \quad (7)$$

$$y_r, y_b \ge 0 \quad (8)$$

$$r(u,v)+r(v,u)+b(u,v)+b(v,u)\le 1, \forall x(u,v)\in E \quad (9)$$

$$r(u,v), b(u,v)\in\{0,1\} \quad (10)$$

The requirement is to find the maximum total bandwidth of the red and blue paths, subject to the following conditions: Condition (1) ensures that the red path is connected. i.e., each node in the path is connected to two nodes, with the exception of s and t nodes, which have to be connected to only one node each. Condition (2) ensures the same for the blue path. Conditions (3), (4) and (5) ensure that, excluding the source and destination nodes, if any node is found in the red path, it is not in the blue path and vice versa. i.e., the path is node-disjoint. Conditions (6) and (7) ensure that bandwidth (yr and yb) of each path has the value of its lowest-bandwidth link. Condition (8) ensures that the bandwidth of each path is not less than zero. Condition (9) ensures that the two paths are also link-disjoint. Finally, condition (10) ensures that r(u, v) and b(u, v) can only take values 0, 1, which means that the link either does not or does exist in the path, respectively.

B. Network Topologies and Setup

We study the performance of our MLBDP algorithm compared to ILP and the MBA heuristic algorithm developed in [17]. The comparative studies were performed on two network topologies: STC backbone network [18], shown in Figure 4, and ARPANET network [19], shown in Figure 5.

For both topologies, we examined networks with maximum possible bandwidth on any link equals to 10, 20, 50, 100, 200, 500, 1000, 2000 and 5000. For each case, we tested the maximum-bandwidth node-disjoint path obtained for each source and destination. The STC backbone network has 35 nodes and 45 links, while the ARPANET network has 20 nodes and 32 links.

The algorithms which have been compared include ILP, as formulated in this section, MBA algorithm, as described in Related Work, and our MLBDP algorithm. Recall that the MBA algorithm aims to find link-disjoint rather than node-disjoint maximum bandwidth paths. However, it can be easily modified to find node-disjoint paths. We made the necessary changes to make it node-disjoint in order to provide a fair comparison. We have used C# programming language for coding our MLBDP algorithm, as well as the MBA algorithm. ILP implementation was also programmed using C# with Microsoft Solver Foundation. Simulations were done on Core 2 Duo computers with about 2GHz speed.

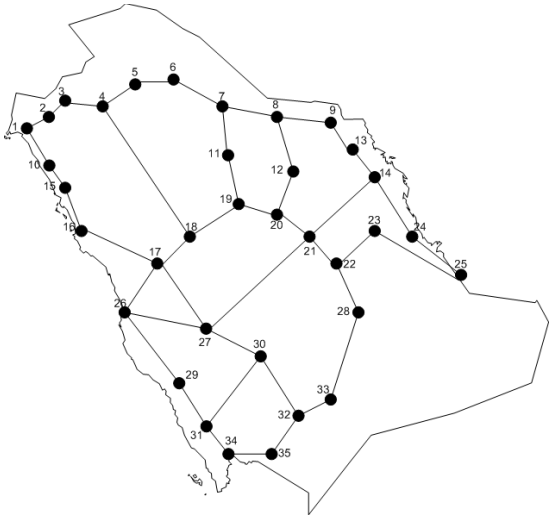


Figure 4. STC backbone network

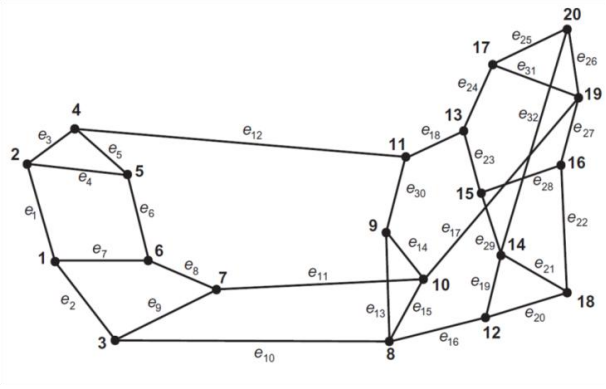


Figure 5. ARPANET network with 20 nodes

C. Simulation Results

Simulation results for ARPANET network and STC backbone network are shown in Table I and Table II, respectively. Both networks have been tested under nine different sets of bandwidth capacities. The bandwidth of each link has been assigned a random value between 1 and a maximum value shown in the first column. Columns 2, 3 and 4 show the number of disjoint paths found, between all source-destination pairs, using MBA, our MLBDP algorithm, and ILP, respectively. For STC backbone network, there are 35 nodes. For each node taken as source, there can be at most 34 maximum-bandwidth node-disjoint paths, one for each destination. i.e., 35 × 34 = 1190 paths. For ARPANET network, this number is calculated as 20 × 19 = 380. We note that both networks are designed so that there is at least a node-disjoint path between each pair of nodes. Columns 5, 6, and 7 show the execution time in milliseconds it took each of the three algorithms to find the maximum-bandwidth node-disjoint paths. It should be noted that the solutions obtained using ILP are considered optimal and can be used to benchmark other algorithms. The maximum bandwidth obtained by the MBA algorithm in each case has been subtracted from the maximum bandwidth obtained by ILP for the same case. The total difference (sum of subtraction results) is shown in column 8 and the average difference (average of subtraction results) is shown in column 9. Columns 10 and 11 show the total and average differences calculated in the same manner between our MLBDP algorithm and ILP.

From the results, it can be observed that ILP execution time is much higher than MBA and MLBDP algorithms. Note that a log scale has been used to make clearer representation in Figure 7 and Figure 9. Although our MLBDP algorithm takes longer time to finish than MBA algorithm, it was able to find significantly higher number of disjoint paths than MBA and thus was able to find higher maximum-bandwidth paths. It can be seen from Figure 6 and Figure 8 that our MLBDP algorithm was able to find all possible disjoint paths obtained by ILP in a much lower execution time (about two orders of magnitude less than ILP). By looking at columns 8 to 11 in Table I and Table II, it can be observed that there is a difference between the maximum bandwidth obtained by MBA and by ILP. The bandwidth difference is higher at lower rows because these rows have higher deviation between the bandwidths available in each link. The reason of MBA failure is the two-step approach used in it. i.e., after finding the first maximum-bandwidth path, all links in this path are removed from the network. The removed links can at worst prevent an existing disjoint path from being found, or at best eliminate a path with higher combined maximum bandwidth. On the other hand, it can be seen that our MLBDP algorithm has a zero difference with ILP in the maximum bandwidth found. That's because the MLBDP algorithm works concurrently on the two disjoint paths, minimizing the probability of missing a better candidate path.

Graphical representations of the results are shown in Figure 6 and Figure 7 for STC backbone network. For ARPANET network, the results are shown in Figure 8 and Figure 9.

TABLE I. SIMULATION RESULTS FOR STC BACKBONE NETWORK

Max BW	Number of Disjoint Paths Found			Execution Time			Difference ILP-MBA		Difference ILP-MLBDP	
	MBA	*MLBDP*	*ILP*	*MBA*	*MLBDP*	*ILP*	*Total*	*Average*	*Total*	*Average*
10	988	1190	1190	1827.1	4103.2	138872.9	334	0.28	0	0
20	934	1190	1190	2354.1	6869.3	173326.9	566	0.48	0	0
50	930	1190	1190	3028	12319.7	205795.7	1124	0.94	0	0
100	930	1190	1190	3146.1	14305.8	202170.5	1938	1.63	0	0
200	929	1190	1190	3161.1	16949.9	192363	3601	3.03	0	0
500	929	1190	1190	3390.1	16927.9	193892	8522	7.16	0	0
1000	929	1190	1190	3427.1	18523	231039.2	16726	14.06	0	0
2000	929	1190	1190	3603.2	18923	220709.6	33194	27.89	0	0
5000	897	1190	1190	3571.2	18646	233137.3	131453	110.46	0	0

TABLE II. SIMULATION RESULTS FOR ARPANET NETWORK

Max BW	Number of Disjoint Paths Found			Execution Time			Difference ILP-MBA		Difference ILP-MLBDP	
	MBA	*MLBDP*	*ILP*	*MBA*	*MLBDP*	*ILP*	*Total*	*Average*	*Total*	*Average*
10	362	380	380	436	429	89235.1	82	0.22	0	0
20	362	380	380	484	888	116026	103	0.27	0	0
50	362	380	380	568	1258	100572	255	0.67	0	0
100	362	380	380	557	1158	118052	468	1.23	0	0
200	357	380	380	588	1352	118377	2278	5.99	0	0
500	357	380	380	728	1417	159979	5811	15.29	0	0
1000	357	380	380	680	1531	138376	11572	30.45	0	0
2000	357	380	380	656	1420	136474	22969	60.44	0	0
5000	357	380	380	566	1490	131096	57262	150.69	0	0

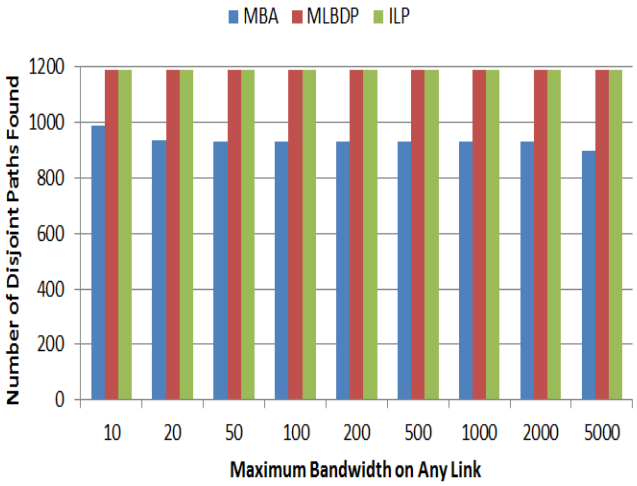


Figure 6. Number of disjoint paths found - STC backbone network

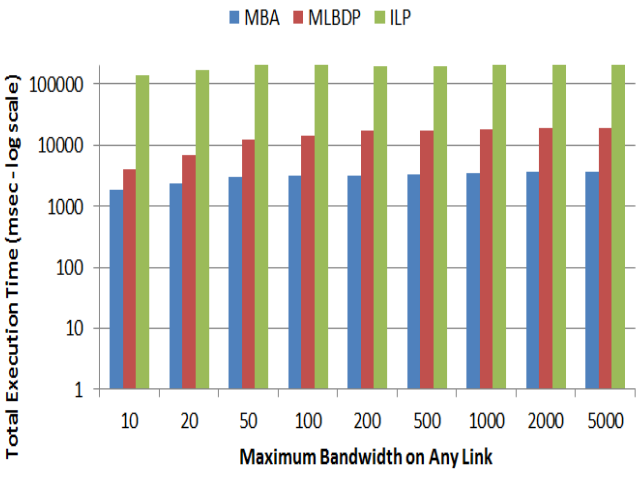


Figure 7. Total execution time - STC backbone network

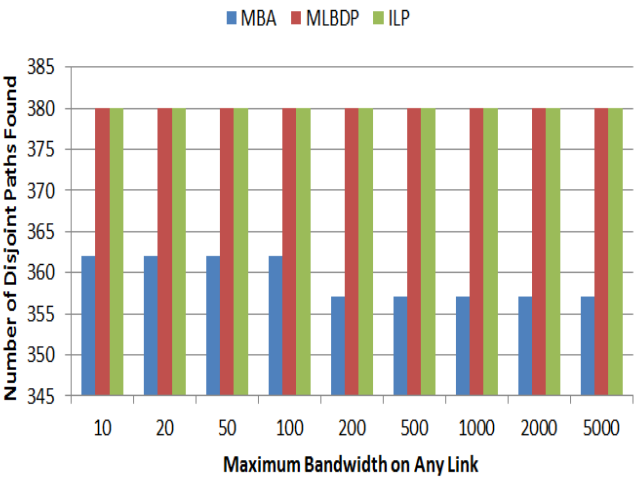


Figure 8. Number of disjoint paths found - ARPANET network

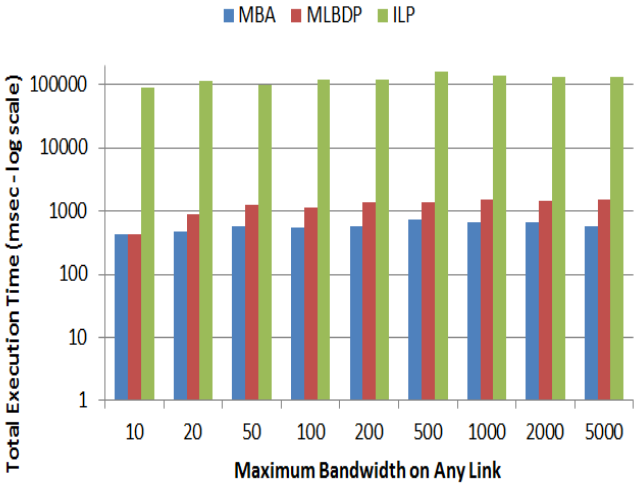


Figure 9. Total execution time - ARPANET network

VII. Analytical Study

From the numerical results in the previous section, we can observe that our MLBDP algorithm finishes in polynomial time. The goal of this section is to provide a general estimation of the execution time.

It has been shown in [20] that Dijkstra algorithm works in:

$$\text{Run}\left(\text{Dijkstra}\right) = O\left(m + n\log n\right) \tag{11}$$

where n is the number of nodes and m is the number of links. The MLBDP algorithm uses a virtual topology with n×n virtual nodes and with 2nm virtual links. To see that the number of virtual links is 2mn, it can be seen from the algorithm description in Figure 2 that the number of links originating from each virtual node (vnode) equals the sum of the number of links originating from both of its constituent (actual) nodes. Let the number of links coming from nodes i and j equal ki and kj, respectively. The number of links originating from vnode[ij] equals ki+kj.

To calculate the total number of links K in the virtual topology, note that each link is connected to two nodes. If we add the number of links originating from each node, every link will be counted twice. Thus, the summation of the number of links should be divided by 2., i.e.

$$K = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(k_i + k_j\right) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}k_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}k_j \tag{12}$$

We can see that $\sum_{j=1}^{n}k_j = 2m$. Similarly $\sum_{i=1}^{n}k_i = 2m$. Substituting in (12) yields:

$$\frac{1}{2}\sum_{i=1}^{n}n\cdot k_i + \frac{1}{2}\sum_{i=1}^{n}2m = \left(\frac{n}{2}\right)\left(2m\right) + \left(2m\right)\left(\frac{1}{2}\right)\left(n\right) = 2mn \tag{13}$$

From (11) and (13) , a single run of the MLBDP algorithm yields a run time of approximately:

$$\begin{aligned}\text{Run}\left(\text{MLBDP}\right) &\approx O\left(2mn + n^2\log n^2\right)\\ &\approx O\left(2mn + 2n^2\log n\right)\end{aligned} \tag{14}$$

Recall from Section 4 that the MLBDP algorithm needs to be executed q times, where q is the number of unique link bandwidths. In the worst case, where each of the m links in the network has a unique bandwidth, q = m. Thus, the worst-case full run of MLBDP algorithm yields a run time of approximately:

$$\begin{aligned}\text{FullRun}\left(\text{MLBDP}\right) &\approx O\left(m\left(2mn + 2n^2\log n\right)\right)\\ &\approx O\left(2m^2n + 2mn^2\log n\right)\end{aligned} \tag{15}$$

VIII. Summary And Conclusions

This paper has presented a new algorithm, Max-Limit Bandwidth Disjoint Path (MLBDP), for finding node-disjoint paths with maximum combined bandwidth. The algorithm works on multiple iterations. At each iteration, the algorithm works on finding two paths concurrently: one with maximum bandwidth and another with bandwidth greater than a certain limit. The limit takes the values of all possible unique link bandwidths. The presented algorithm uses a modified Dijkstra algorithm and a virtual network topology. A performance comparison has been done for the MLBDP algorithm to a modern heuristic, MBA algorithm, and to the optimal solution using Integer Linear Programming (ILP). The simulation studies have shown that the MLBDP algorithm was able to obtain results identical to the ILP solution at a significantly lower execution time. In addition, the MLBDP avoids the MBA problem of missing valid disjoint paths because it works on the two disjoin paths concurrently. Thus, despite the slightly additional execution time, the MLBDP algorithm offers an overall better performance over the MBA algorithm. Future enhancements of the presented MLBDP algorithm can be done to reduce the execution time. Also, the algorithm can be modified to find maximum-bandwidth link-disjoint paths.

Acknowledgment

The author would like to thank the Research Center of the College of Computer and Information Sciences, King Saud University, for their support.

References

[1] G. Xue, A. Sen, W. Zhang, J. Tang, and K. Thulasiraman, “Finding a Path Subject to Many Additive QoS Constraints,” Networking, IEEE/ACM Transactions on, vol. 15, no. 1, pp. 201-211, 2007.

[2] S. Bistarelli, U. Montanari, F. Rossi, and F. Santini, “Unicast and multicast QoS routing with soft-constraint logic programming,” ACM Trans. Comput. Logic, vol. 12, no. 1, pp. 5:1–5:48, Nov. 2010.

[3] G. Murugesan, J. Jebarani, and A. M. Natarajan, “Adaptive Granularity Algorithm for Effective Distributed Load Balancing and Implementation in Multiprotocol Label Switching Networks,” in International Conference on Advanced Computing and Communications, 2007. ADCOM 2007, 2007, pp. 626-633.

[4] C. Li, P. Li, and T. Mohammed, “An Optimal MPLS-TE Solution to Route Selection and Redistribution on Congested Networks,” in International Conference on Networking, Architecture, and Storage, 2007. NAS 2007, 2007, pp. 69-76.

[5] P. Van Mieghem and F. A. Kuipers, “On the Complexity of QoS Routing,” COMPUTER COMMUNICATIONS, vol. 26, p. 376--387, 2003.

[6] F. A. Kuipers and P. F. . VanMieghem, “Conditions That Impact the Complexity of QoS Routing,” IEEE/ACM Transactions on Networking, vol. 13, no. 4, pp. 717- 730, Aug. 2005.

[7] R. Bhandari, “Optimal diverse routing in telecommunication fiber networks,” in , 13th Proceedings IEEE INFOCOM ’94. Networking for Global Communications, 1994, pp. 1498-1508 vol.3.

[8] W. Liang, “Robust routing in wide-area WDM networks,” in Parallel and Distributed Processing Symposium., Proceedings 15th International, 2001.

[9] B. H. Shen, B. Hao, and A. Sen, “On multipath routing using widest pair of disjoint paths,” in 2004 Workshop on High Performance Switching and Routing, 2004. HPSR, 2004, pp. 134- 140.

[10] S. Sahni, N. Rao, S. Ranka, Y. Li, E.-S. Jung, and N. Kamath, “Bandwidth Scheduling and Path Computation Algorithms for Connection-Oriented Networks,” in Sixth International Conference on Networking, 2007. ICN ’07, 2007, pp. 47-47.

[11] J. Suurballe, “Disjoint paths in a network,” Networks, vol. 4, no. 2, pp. 125-145, 1974.

[12] J. Suurballe and R. Tarjan, “A quick method for finding shortest pairs of disjoint paths,” Networks, vol. 14, no. 2, pp. 325-336, 1984.

[13] A. Sen, Bin Hao, Bao Hong Shen, Ling Zhou, and S. Ganguly, “On maximum available bandwidth through disjoint paths,” in 2005 Workshop on High Performance Switching and Routing, 2005. HPSR, 2005, pp. 34- 38.

[14] T. D. Am, S. Tsukiyama, I. Shirakawa, and H. Ozaki, "An algorithm for generating all the paths between two vertices in a digraph and its application," Trans. Information Processing Soc. Japan, vol. 16, no. 9, pp. 774–780, 1975.

[15] H. Leng, M. Liang, J. Song, Z. Xie, and J. Zhang, "Routing on Shortest Pair of Disjoint Paths with Bandwidth Guaranteed," in Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing, 2009. DASC '09, 2009, pp. 557-561.

[16] G. N. Frederickson, "An optimal algorithm for selection in a min-heap," Inf. Comput., vol. 104, no. 2, pp. 197–214, Jun. 1993.

[17] R. C. Loh, S. Soh, and M. Lazarescu, "Maximizing Bandwidth Using Disjoint Paths," in 2010 24th IEEE International Conference on Advanced Information Networking and Applications (AINA), 2010, pp. 304-311.

[18] "STC - Wholesale Network Infrastructure." [Online]. Available: http://www.stc.com.sa/cws/portal/en/business/bus-wholesale/stc-lnd-whlslr-news_and_acti/stc-lnd-whlslr-infcap. [Accessed: 09-Dec-2011].

[19] R. Andersen, F. Chung, A. Sen, and G. Xue, "On Disjoint Path Pairs with Wavelength Continuity Constraint in WDM Networks," IN WDM NETWORKS; IEEE INFOCOM'2004, vol. 2, pp. 524-535, 2004.

[20] J. Sneyers, T. Schrijvers, and B. Demoen, "Dijkstra's algorithm with Fibonacci heaps: An executable description," IN CHR. IN 20TH WORKSHOP ON LOGIC PROGRAMMING (WLP'06, p. 182--191, 2006.

AUTHORS PROFILE

Mostafa H. Dahshan has received his B.S. degree in Computer Engineering from Cairo University, Egypt in 1999. He received his M.S. in Telecomm. Systems and Ph.D. in Electrical and Computer Engineering from the University of Oklahoma, USA in 2002 and 2006, respectively. He is currently an Assistant Professor of Computer Engineering at the College of Computer and Information Sciences, King Saud University, Saudi Arabia. His current research interests include Network Protocols, Performance, Reliability and Security.

Message Segmentation to Enhance the Security of LSB Image Steganography

Dr. Mohammed Abbas Fadhil Al-Husainy

Department of Multimedia Systems,
Faculty of Sciences and Information Technology,
Al-Zaytoonah University of Jordan.
Amman, Jordan

***Abstract*—Classic Least Significant Bit (LSB) steganography technique is the most used technique to hide secret information in the least significant bit of the pixels in the stego-image. This paper proposed a technique by splitting the secret message into set of segments, that have same length (number of characters), and find the best LSBs of pixels in the stego-image that are matched to each segment. The main goal of this technique is to minimize the number of LSBs that are changed when substituting them with the bits of characters in the secret message. This will lead to decrease the distortion (noise) that is occurred in the pixels of the stego-image and as result increase the immunity of the stego-image against the visual attack. The experiment shows that the proposed technique gives good enhancement to the Classic Least Significant Bit (LSB) technique.**

Keywords—Security; Distortion; Embedding; Substitution.

I. INTRODUCTION

Steganography is one of many techniques that are used to hide secret information to prevent any attackers to make damage in this information or use it in illegal form. Steganography can be defined as the technique used to embed data or other secret information inside some other object commonly referred to as cover, by changing its properties. The purpose of steganography is to set up a secret communication path between two parties such that any person in the middle cannot detect its existence; the attacker should not gain any information about the embedded data by simply looking at cover file or stego file. Steganography is the art of hiding information in ways that prevent the detection of hidden messages. Steganography, derived from Greek, literally means "covered writing." It includes a vast array of secret communications methods that conceal the message's very existence. These methods include invisible inks, microdots, character arrangement, digital signatures, covert channels, and spread spectrum [1, 2].

The basic model of steganography uses a cover object (any object that can be used to hold secret information inside), the secret message (the secret information that is to be sent to some remote place secretly), a stego key that is used to encode the secret message to make its detection difficult and a steganography algorithm/technique (the procedure to hide secret message inside cover object). The outcome of the process is the stego object which is the object that has the secret message hidden inside. This stego object is sent to the receiver where receiver will get the secret data out from the stego image by applying decoding algorithm/technique [1].

Recently, steganography is implemented by using digital media. Secret message is embedded inside digital cover media like text, images, audio, video or protocols depending upon the requirement and choice of the sender. Compared with the other types of steganography, the image steganography is most widely used. The reason behind the popularity of image steganography is the large amount of redundant information present in the images that can be easily altered to hide secret messages inside them, and because it can take advantage of the limited power of the human visual system (HVS). With the continued growth of strong graphics power in computer and the research being put into image based steganography, this field will continue to grow at a very rapid pace [1, 3, 4, 5].

Steganography has a wide range of applications. The major application of steganography is for secret data communication. Covert channels in TCP/IP involve masking identification information in the TCP/IP headers to hide the true identity of one or more systems. Cryptography is also used for the same purpose but steganography is more widely used technique as it hides the existence of secret data. Another application of steganography is feature tagging. Captions, annotations, time stamps, and other descriptive elements can be embedded inside an image, such as the names of individuals in a photo or locations in a map [1, 2, 5, 6].

Steganography can be also used to combine explanatory information with an image (like doctor's notes accompanying an X-ray). Steganography is used by some modern printers, including HP and Xerox brand color laser printers. Tiny yellow dots are added to each page. The dots are barely visible and contain encoded printer serial numbers, as well as date and time stamps. The application list of image steganography is very long [1, 6].

The Steganography technique is the perfect supplement for encryption that allows a user to hide large amounts of information within an image. Thus, it is often used in conjunction with cryptography so that the information is doubly protected; first it is encrypted and then hidden so that an adversary has to first find the hidden information before decryption take place [7, 8, 9, 10, 26(11)]. The problem with cryptography is that the encrypted message is obvious. This means that anyone who observes an encrypted message in

transit can reasonably assume that the sender of the message does not want it to be read by casual observers. This makes it possible to deduce the valuable information. Thus, if the sensitive information will be transmitted over unsecured channel such as the internet, steganography technique can be used to provide an additional protection on a secret message [2].

A good technique of image steganography aims at three aspects. First one is capacity (the maximum data that can be stored inside cover image). Second one is the imperceptibility (the visual quality of stego-image after data hiding) and the last is robustness [7].

The idea in this paper is that when substitute the LSB of the pixels (in the stego-image) with the bits of all characters (in the secret message) as one segment, this will result a large number of changes that are happen in LSB of pixels. It is normally come from the truth that it is rarely find a best match between very long sequence of bits of all characters in the secret message and the LSB of the pixels in the stego-image. A message segmentation LSB was proposed in this paper to overcome this problem by splitting the secret message into set of segments of same length (same number of characters). And try to find the best match between the bits of the characters in each segment and the LSB of different sequences of pixels in the stego-image. When the proposed technique split the long secret message into number of small segments, this will lead to increase the probability of finding best matching between the bits of the characters in the secret message and the LSB of the pixels in the stego-image. The best match between bits will decrease the number of LSB of the pixels that are changed when replace the bits of characters in the secret message in it. As a result of that, the distortion/noise that will appear in the pixels of the stego-image will be decrease and the immunity of the stego-image against the attack by human visual system (HVS) becomes strong.

II. Related Works

When hiding information inside images usually Least Significant Bit (LSB) method is used. In the LSB method the 8th bit of every byte of the carrier file is substituted by one bit of every bit of the secret information [12]. The LSB method usually does not increase the file size, but depending on the size of the information that is to be hidden inside the file, the file can become noticeably distorted.

Ross J. Anderson and Fabien A.P. Petitcolas argued that every steganographic approach will have its limitations; they proposed an information theoretic approach using Shannon's theory for perfect secrecy [13]. In the methods that are proposed by H. Motameni and his colleague's one can embed at the dark corners of an image [14]. One can also embed the secret information in frequency domain by using Discrete Wavelet Transform method [15]. In this method the embedding should be done at high frequency coefficients. P. Mohan Kumar and D. Roopa suggested that one can apply block matching procedure to search the highest similarity block for each block of the secret image and embed in LSBs of the cover image [16]. Mohammed A.F. AlHusainy employed different strategy in image steganography art by mapping the pixels of image to English letters and special characters [17]. Lisa M Marvel and Charles G Boncelet proposed to hide at the inherent noise places [18]. Ran-Zan Wang and Yeh-shun Chen also did the two way block matching for image in image steganography [19]. But this approach is suspicious to the hackers. Xinpeng Zhang and his colleagues proposed an approach called "multibit assignment steganography for palette images", in which each gregarious color that possesses close neighboring color in the palette is exploited to represent several secret bits [20]. In reference [21] authors have discussed a double substitution algorithm for encrypting at sender and decrypting at receiver and the embedding process was at 7th and 8th bit positions alternatively. In [22] an image steganography with palette based images is suggested. The method is based on a palette modification scheme, which can iteratively embed one message bit into each pixel in a palette based image. In each iteration, both the cost of removing an entry color in a palette and the benefit of generating a new one to replace it are calculated. If the maximal benefit exceeds the minimal cost, an entry color is replaced. It is found that the fundamental statistics of natural images are altered by the hidden non-natural information [23]. But if we do not touch the bytes those carry the image features and embed in the other bytes then the problem can be solved. As LSB embedding is very common, many steganalysis tools are available for it [24]. So LSB embedding is no more secured now-a-days. So, new embedding techniques are to be welcomed to the steganographic world. Due to the large number of steganographic tools available over the internet, a particular threat exists when criminals use steganography to conceal their activities with in digital images in cyber space. Reference [25] presents two JPEG steganographic methods using Quantization Index Modulation (QIM) in the Discrete Cosine Transform (DCT) domain. The two methods approximately preserve the histogram of quantized DCT coefficients, aiming at secure steganography against histogram-based attacks. Sukhpreet Kaur and Sumeet Kaur in [26] developed a technique for hiding text using image steganography that use 7 bits per pixel as a full capacity of the cover image to hide data and still no visual changes in the stego image.

III. Classic-Lsb Image Steganography Technique

The Least Significant Bit (LSB) steganography technique works by representing each character (byte) of the secret message as a set of 8-bits (where 1 byte ≡ 8 bits). And then hide/replace the bits of the characters in the least significant bit of the pixels in the stego-image. If the secret message has n characters, then LSB technique need at least (n*8) pixels in the stego-image to hid the bits of the n characters.

By substitute the LSB of each pixel in the stego-image with one bit (from the 8-bits) of each character in the secret message, the substitution operation will cause some distortion/noise in the stego-image. By using Human Visual System (HVS), the attackers may doubt that the stego-image contain a secret information in it. In general, whenever the length of the secret message (number of characters) is long, then the noise in the stego-image probably will increase as a result. This will make restriction to hide a very long message in a small stego-image. Therefore, we will tend to choose a short message to hide it in a large stego-image to minimize the noise that is happen in the pixels of the stego-image and to put aside

the doubt about containing the stego-image any secret information.

Also, when an attacker success know that the stego-image contains a secret message, it is easy to get this message by recompose the secret message from the LSB of the pixels in the stego-image.

IV. The Proposed LSB Image Steganography Technique

The message segmentation LSB technique is suggested here to enhance the performance of the Classic-LSB technique by supporting it through three strong points:

- Decrease the distortion/noise that will be appearing in the pixels of the stego-image.
- Increase the capability of hiding very long secret message in a small stego-image.
- Increase the immunity of the stego-image against the attacks of Human Visual System (HVS).

In the following paragraphs, the detail explanation of the operations that are doing in the proposed technique will be given. Two definitions used in this technique for secret message and stego-image are listed below:

A secret message is an English message might be contains alphabetic letters (‘a’…’z’) or numbers (‘0’…’9’) or any special symbols like: ('space character', '.', ',', '(', ')').

A stego-image, for the purpose of testing, a candidate image to be used in this work is a bitmap images (.bmp) type. In general, each file of type (.bmp) is consisting of a header part which is containing much information like (Width and Height of the image, number Palette, number of bits for each pixel) followed by the data of the bitmap image pixels. The pixels of each image represent as a two dimensional list, but the proposed technique treat the pixels of the image as a one dimensional list of bytes, (where each byte has a value between (0…255)), by reading the bytes of the two dimensional image row by row and stores them as a one dimensional list.

Before listing the steps of the algorithm that describe the operations of the proposed technique, some data structures used in the algorithm are defined below:

1) MessageB: is a list that contains a binary representation (bits) of all characters in the secret message. The number of elements (size) of this list is (n*8), where n is the number of characters in the secret message.

2) ImageB: is a list of the Least Significant Bit (LSB) of all pixels in the stego-image. The number of elements (size) of this list is (m), where m is the size of the image and its equal (Width × Height × Palette).

3) SegmentLength: is a positive integer number between (2 … (n*8)/2) which represents the length of each segment (number of bits) in the SegmentList.

4) SegmentsList: is a list of segments that is created from the MessageB by splitting it to k segments, where k = (n*8) / SegmentLength. And each segment has number of bits equal SegmentLength.

5) *SegmentIndex:* is a list of indices, each index represents the first index of a sequence of bits in ImageB which have a best match with the bits for one of the segments in SegmentsList. We must note that there is no overlapping between the matched bits sequences in this technique.

Algorithm:

// Hiding Operation

Step1: Calculate the ***TotalSize*** (in byte) that is required to store:

(1) Length of secret message (number of character)

(2) ***SegmentLength***

(3) Size of ***SegmentList***

Step2: Store the bits representation of the above three information in the Least Significant Bit (LSB) at the start of the ***ImageB*** list (from **bit #1** to **bit #(*TotalSize**8)**).

Step3: For ***i* = 1** To **((*n**8) / *SegmentLength*)**

{

For ***j* = ((*TotalSize**8)+1)** To ***m***

{

***x* = 1**

***BestMatch* = 0**

***BestIndex* = -1**

For ***w* = *j*** To **(*j*+ *SegmentLength*)**

{

Find the number of matched bits ***MBits*** in ***Segment*[*i*][*x*]** with the bits of ***ImageB*[*w*]**

***x* = *x*+1**

}

If (***MBits*** > ***BestMatch***)

{

BestMatch* = *MBits

BestIndex* = *j

}

}

SegmentIndex*[*i*] = *BestIndex

Substitute the bits of ***Segment*[*i*]** instead of the bits in ***ImageB*** starting at ***BestIndex***

}

// Extracting Operation

Step1: Read from the stego-image the information that is stored in the first **(*TotalSize**8)** LSB of the pixels.

Step2: Reconstruct the segments of the secret message by using the extracted information in Step1.

Step3: Reassembling the all the segments that are constructed in Step2 to regenerate the characters of the secret message.

V. Experimental Results And Discussion

The performance of the proposed Segmented-LSB image steganography technique has been tested by using both the Classic-LSB and the proposed LSB to hide some messages in different (.bmp) images and record the results to enable the reader to makes a good comparison, in the performance, between these two techniques. Table I shows the stego-images, of different sizes, that are used in the experiments. Table II summarizes the recorded results from the experiments using SegmentLength = 10.

To clarify the effect of SegmentLength on the performance of the proposed Segmented-LSB image steganography technique, different selected values of the SegmentLength used on the above experiments. Fig. 1 shows the effect of the SegmentLength on (a) Number of LSB changed, (b) Signal to Noise Ratio (SNR) of the stego-image, (c) Time of hiding operation.

The required programs to implement the Classic-LSB and the proposed LSB techniques written by using C++ programming language and executing them on a computer system are of 2.53GHz processor with 4.0 GB memory and Microsoft Windows 7 operating system.

From Table I, we note that the proposed LSB decrease the number of LSB that are changed in the stego-image when it compares with the Classic-LSB, this certainly enhance the SNR of the stego-image. The main challenge of the proposed LSB is the time of the hiding operation, but this comes from performing the exhaustive search to find the best matching between the bits in each segment with all non-overlapped bits sequences in the ImageB list.

When we see the three parts of Fig. 1, we can note that the proposed LSB produce a stable performance when the SegmentLength change:

- When increase the SegmentLength the number of LSB that are changed will increase and vice versa. This is because when the SegmentLength be large the possibility of finding best match between bits becomes less.
- When increase the SegmentLength the SNR will decrease and vice versa. This is because the value of SNR of the stego-image is proportional with the number of LSB that are changed in the pixels of the stego-image.
- The time of the hiding operation of each image was increase/decrease with few changes. It stays suitable when the size of the stego-image is small, but it will be long when the size of the stego-image becomes large. This is because the search time for best matching becomes huge when we using a stego-image of large size.

TABLE I. Stego-Images (.Bmp) Of Size (Width × Height × Palette)

Butterfly (128×128×3)

Garden (160×121×3)

Girls (300×225×3)

TABLE II. Recorded Results of Performance Experiments

	Classic-LSB Technique			Proposed LSB Technique		
Stego-Image	***Butterfly***	***Garden***	***Girls***	***Butterfly***	***Garden***	***Girls***
Length of Secret Message (Characters)	500	1000	2000	500	1000	2000
Number of LSB Changed	1972	4001	8081	1508	2987	5760
Signal to Noise Ratio (SNR) of the Stego-Image	51.950	50.107	51.890	53.115	51.377	53.360
Time of Hiding Operation (Second)	0.047	0.140	0.421	3.76	7.005	60.372
Time of Extracting Operation (Second)	0.110	0.125	0.421	0.109	0.124	0.421

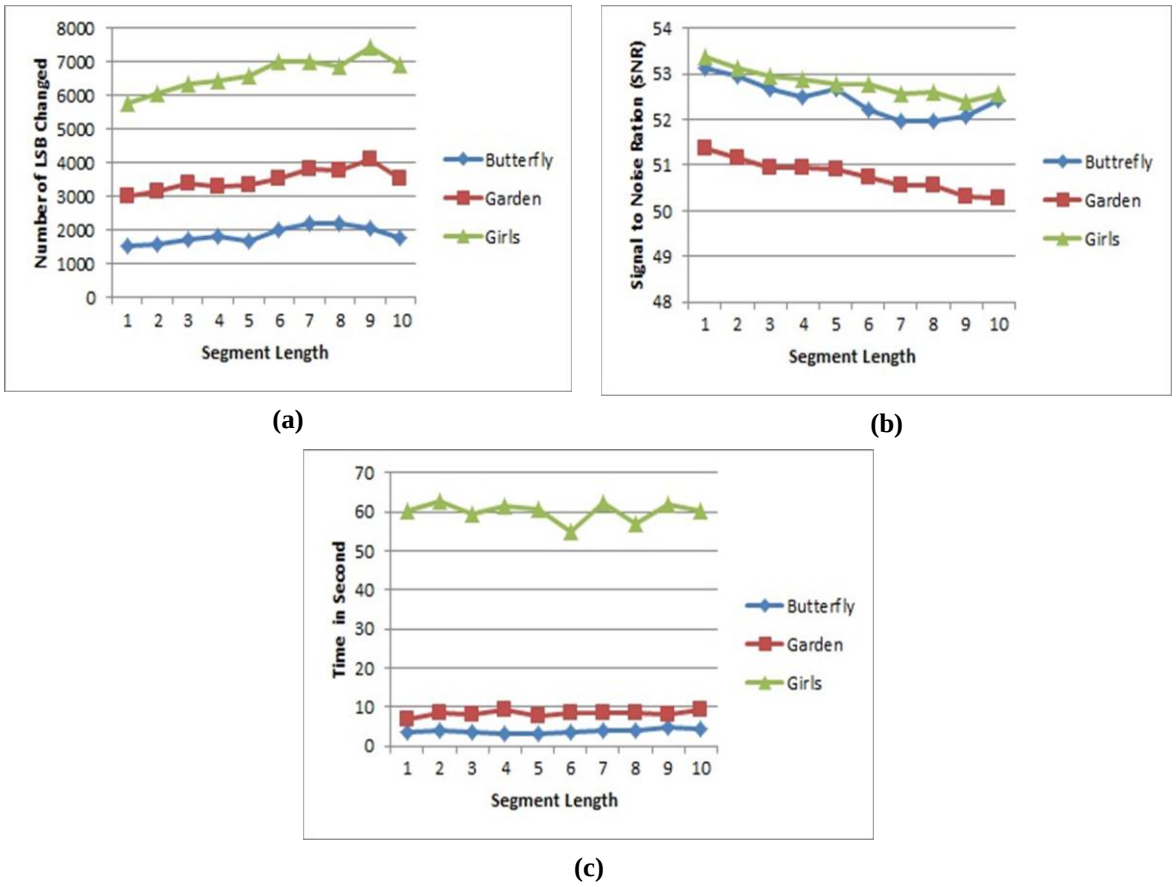


Figure 1. The effect of the SegmentLength on (a) Number of LSB changed, (b) Signal to Noise Ratio (SNR) of the stego-image, (c) Time of hiding operation.

VI. CONCLUSION

The idea to enhance the performance of the Classic-LSB image steganography technique was present in this paper. The message segmentation LSB image steganography technique was suggested here by splitting the long secret message into number of short segments. Then hide these short segments in different parts of the best matched LSB in the pixels of the stego-image. The main goal behind this suggested technique is to decrease the number of LSB that are changed of the pixels in the stego-image and as a result increase the immunity of the stego-image against the attack by human visual system (HVS). The recorded results from the experiments showed that the proposed LSB image steganography technique success in increase the security of the secret message that is hid in the stego-image by decreasing the number of LSB that are changed in the pixels of the stego-image.

The challenge point of the proposed LSB is in the long time of the hiding operation that is spend during the exhaustive search to find the best matching when using a large size stego-image. I

n the next work, we will try to minimize the effect of this weak point on the performance of the proposed LSB. But in spite of this point, the Segmented-LSB still can be used instead of the Classic-LSB to satisfy more security for the secret message.

REFERENCES

[1] Cheddad, J. Condell, K. Curran, & P. Kevitt, (2010). Digital image steganography- survey and analysis of current methods. Signal Processing, 90, 727–752. doi: http://10.1016/j.sigpro.2009.08.010

[2] Adnan Gutub, Ayed Al-Qahtani, & Abdulaziz Tabakh. (2009). Triple-A: secure RGB image steganography based on randomization. AICCSA, IEEE/ACS International Conference on Computer Systems and Applications, Rabat, Morocco, 400-403, doi: http://doi.ieeecomputersociety.org/10.1109/AICCSA.2009.5069356

[3] Kaur, R. Dhir, & G. Sikka. (2009). A new image steganography based on first component alteration technique. International Journal of Computer Science and Information Security (IJCSIS), 6, 53-56. http://arxiv.org/ftp/arxiv/papers/1001/1001.1972.pdf

[4] Alvaro Martin, Guillermo Sapiro, & Gadiel Seroussi. (2005). Is Steganography Natural. IEEE Transactions on Image Processing, 14(12), 2040-2050. doi: 10.1109/TIP.2005.859370

[5] Bhattacharyya, A. Roy, P. Roy, & T. Kim. (2009). Receiver compatible data hiding in color image. International Journal of Advanced Science and Technology, 6, 15-24. http://www.sersc.org/journals/IJAST/vol6/2.pdf

[6] EE. Kisik Chang, J. Changho, & L. Sangjin. (2004). High Quality Perceptual Steganographic Techniques. Springer. 2939, 518-531. doi: 10.1007/978-3-540-24624-4_42, http://www.springerlink.com/content/c6guuj5xnyy4wj3c/

[7] C. Kessler. (2001). Steganography: Hiding Data within Data. An edited version of this paper with the title "Hiding Data in Data". Windows & .NET Magazine. [Online] Available: http://www.garykessler.net/library/steganography.html (October 4,2011)

[8] Gandharba Swain, & S.K.lenka. (2010). Steganography-Using a Double Substitution Cipher. International Journal of Wireless Communications and Networking. 2(1), 35-39. ISSN: 0975-7163. http://www.serialspublications.com/journals1.asp?jid=436&jtype

[9] Hideki Noda, Michiharu Nimi, & Eiji Kawaguchi. (2006). High-performance JPEG steganography using Quantization index modulation in DCT domain. Pattern Recognition Letters, 27, 455-46. http://ds.lib.kyutech.ac.jp/dspace/bitstream/10228/450/1/repository6.pdf

[10] Kathryn (2005). A Java Steganography Tool. http://diit.sourceforge.net/files/Proposal.pdf

[11] Gandharba Swain, Dodda Ravi Kumar, Anita Pradhan, Saroj Kumar Lenka, (2010). A Technique for Secure Communication Using Message Dependent Steganography. Special Issue of IJCCT Vol. 2 Issue 2, 3, 4; 2010 for International Conference [ICCT-2010], 3rd - 5th December. http://interscience.in/SpIss_ijcct_icct2010vol2_no234/32_EC31.pdf

[12] Motameni, M.Norouzi, M.Jahandar, & A. Hatami. (2007). Labeling method in Steganography. Proceedings of world academy of science, engineering and technology, 24, 349-354. ISSN 1307-6884. http://www.waset.org/journals/waset/v30/v30-66.pdf

[13] Zhang, & H. Tang. (2007). A novel image steganography algorithm against statistical analysis. Proceeding of the IEEE, 19, 3884-3888. doi: 10.1109/ICMLC.2007.4370824

[14] Lisa M. Marvel, & Charles G. Boncelet. (1999). Spread Spectrum Image Steganography. IEEE Transactions on Image Processing, 8(8), 1075-1083. doi: 10.1109/83.777088

[15] Mei-Yi Wu, Yu-Kun Ho, & Jia-Hong Lee. (2004). An iterative method of palette-based image steganography. Pattern Recognition Letters, 25, 301-309. doi: 10.1016/j.patrec.2003.10.013

[16] Mohammed A.F Al Husainy. (2009). Image Steganography by mapping Pixels to letters. Journal of Computer Science, 5(1), 33-38. ISSN 1549-3636. doi: 10.3844/jcssp.2009.33.38, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.165.7818&rep=rep1&type=pdf

[17] Mohammad Ali Bani Younes, & Aman Jantan. (2008). A New Steganography Approach for Image Encryption Exchange by using the LSB insertion. IJCSNS International Journal of Computer Science and Network Security, 8(6), 247-254. http://paper.ijcsns.org/07_book/200806/20080634.pdf

[18] M.T. Parvez , & A. Gutub. (2008). RGB intensity based variable-bits image steganography. APSCC 2008 –Proceedings of 3rd IEEE Asia-Pacific Services Computing Conference, Yilan, Taiwan, doi: 10.1109/APSCC.2008.105

[19] N.F. Johnson, & J. Suhil. (2006). Exploring Steganography:Seeing the Unseen. Computing Practices. http://www.jjtc.com/pub/r2026.pdf

[20] P.Mohan Kumar, & D.Roopa (2007). An Image Steganography Framework with Improved Tamper Proofing. Asian Journal of Information Technology, 6(10), 1023-1029. ISSN: 1682-3915. http://medwelljournals.com/abstract/?doi=ajit.2007.1023.1029

[21] Po Yuch Chen, & Hung Ju Lin. (2006). A DWT Based Approach for Image Steganography. International journal of Applied Science and Engineering, 4(3), 275-290. http://www.cyut.edu.tw/~ijase/2006/4-3(Microsoft%20Word%20-%20010-009-6).pdf

[22] Ran-Zan Wang, & Yeh-Shun Chen. (2006). High Payload Image Steganography Using Two-Way Block Matching. IEEE Signal Processing letters, 13(3), 161-164. doi: 10.1109/LSP.2005.862603

[23] Ross J. Anderson, & Fabian A.P. Petitcolas. (1998). On The Limits of steganography. IEEE Journal of selected Areas in communication, 16(4), 474-481. Special Issue on Copyright and Privacy protection. ISSN 0733-8716. http://www.cl.cam.ac.uk/~rja14/Papers/jsac98-limsteg.pdf

[24] Sorina Dumitrescu, & Xiaolin (2005). A New Framework of LSB Steganalysis of Digital Media. IEEE Transactions on Signal Processing, 53(10), 3936-3947. doi: 10.1109/TSP.2005.855078

[25] Xinpeng Zhang, Shuozhong Wang, & Zhenyu Zhou. (2008). Multibit Assignment Steganography in Palette Images. IEEE Signal Processing Transactions, 15, 553-556. doi: 10.1109/LSP.2008.2001117

[26] Sukhpreet Kaur, Sumeet Kaur (2010). A Novel Approach for Hiding Text Using Image Steganography. (IJCSIS) International Journal of Computer Science and Information Security, 8(7), October. http://www.scribd.com/doc/40763180/A-Novel-Approach-for-Hiding-Text-Using-Image-Steganography

AUTHORS PROFILE

Mohammed Abbas Fadhil Al-Husainy was born in Mosul, Iraq, in January 1973. He received the M.Sc. and Ph.D. degrees in 1996 and 2002, respectively. From 1997 to 2002, he was a lecturer in the Department of Computer Science, Al-Hadba University of Mosul. Since 2002 he has been an assistant professor in the Departments: Computer Science and Multimedia Systems, Faculty of Science and Information Technology, Al-Zaytoonah University of Jordan. He lectures in the areas of microprocessors, data structures, algorithm design and analysis, digital design systems, operating systems, cryptography, computer organization, programming languages. His research interests are in the broad field of algorithm design, including multi-media data processing, scheduling algorithms, Information Security and cryptography algorithms.

Mobile Learning Environment System (MLES): The Case of Android-based Learning Application on Undergraduates' Learning

Hafizul Fahri Hanafi
Computing Department Sultan Idris Education University,
Perak, Malaysia

Khairulanuar Samsudin
Computing Department Sultan Idris Education University,
Perak, Malaysia

***Abstract*—Of late, mobile technology has introduced new, novel environment that can be capitalized to further enrich the teaching and learning process in classrooms. Taking cognizance of this promising setting, a study was undertaken to investigate the impact of such an environment enabled by android platform on the learning process among undergraduates of Sultan Idris Education University, Malaysia; in particular, this paper discusses critical aspects of the design and implementation of the android learning system. Data were collected through a survey involving 56 respondents, and these data were analyzed by using SPSS 12.0. Findings showed that the respondents were very receptive to the interactivity, accessibility, and convenience of the system, but they were quite frustrated with the occasional interruptions due to internet connectivity problems. Overall, the mobile learning system can be utilized as an inexpensive but potent learning tool that complements undergraduates' learning process.**

Keywords-mobile learning; android learning; teaching and learning.

I. INTRODUCTION

Mobile technology has entered into the mainstream society, affecting the lives of many in recent years. This novel technology is slowly making its presence in the educational realm, which accords many opportunities to the learning and training. Its emergence in the educational world seems timely given the nature of today's learning requirements: wider, fast access to learning materials and persistent needs for prompt communication. Thus, learning institutions must seek every avenue for improvements to cope with new demands of teaching and learning process. Not surprisingly, new and emerging technologies are being passionately sought after by many institutions to provide better learning environments to their stakeholders, namely students and educators. One fine example of the adoption of new technologies is e-learning systems that have radically transformed learning–from being confined within the school walls to borderless landscape, empowering many trainees, students, pupils and other to learn with more academic rigor.

With these formidable learning systems, more and more people can now seek informal education, irrespective of the academic background. Another benefit of such tools is that learning cost incurred unto students is drastically reduced as independent, self-paced learning can be done outside the schools and campuses. Based on these backdrops, it is eagerly anticipated that a new learning approach called mobile learning will be the next major enabler in this decade [1] that will take learning to another level as learning can be literally be conducted on our palms thorough wireless technologies [2]. Given the mass technological consumption of this new technology, a new learning paradigm will dawn over the academic horizon, bringing in new learning opportunities to all.

II. BACKGROUND AND RELATED WORKS

Mobile Learning or M-Learning is a type of e-learning that delivers educational contents and learning support materials through wireless communication devices [3]. Likewise, Traxler [4] describes mobile learning as a personalized, connected, and interactive use of handheld computers in classrooms, in collaborative learning during fieldwork, and in counseling and guidance. All these new learning activities are now possible through M-Learning which is empowered by recent advancements in mobile technology operating systems, notably the ubiquitous android platform. Android technology enables users to communicate with anyone at any time and place almost instantaneously transcending many barriers.

As expected, mobile phones based on android platform have become an indispensable communication device for many people, particularly in younger segments of the population, such as school students. Android is an open source mobile operating system that has been supported by Google Corporation, the world leading search Engine Company. One major reason for the pervasive adoption of android in the mobile market is that mobile applications developed through android development technology is more efficient and effective compared to the other technologies, such as mobile Window or Symbian operating systems, producing fast, user friendly and appealing applications.

As application system files running on android are freely distributed in its Application Market, which is easily accessed over the internet, more and more people are attracted to use this operating system for their mobile devices. Moreover, android-based applications can be run on virtually any personal computers through the android emulator; and this capability promotes the growth of android market globally, leaving behind many rivals in its trail.

III. LEARNING FRAMEWORK

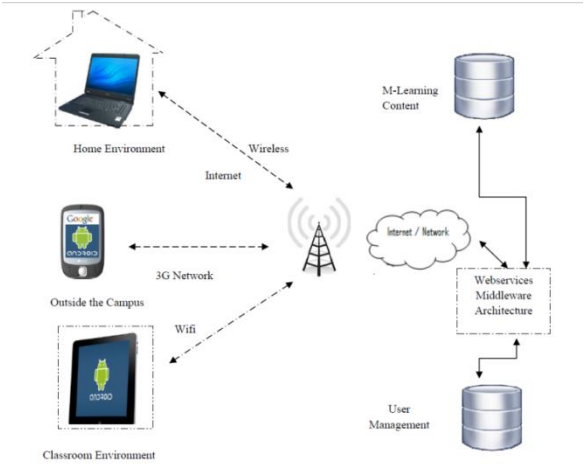


Figure 1. Mobile learning Application Learning Environment [3]

Mobile learning is developed by using multi platforms, languages, and technologies. Thus, learning can be carried out anywhere, anytime for as long as an institution's networking system can gain access to the wireless coverage [5] as illustrated in Figure 1. In this regard, android technology can help realize a mobile learning environment based on the network architecture shown with students gaining fast access to learning contents and materials of their studies by using their mobile phones.

This approach of learning is highly receptive to students as they are more likely to seek and use learning contents via mobile services rather than to find proprietary courseware that is not easily accessed. Propelled by the growing market of smart phones, M-learning is becoming more acceptable in teaching and learning process as these mobile devices are smart as they claimed to be – customizing their contents according to users' specific needs [6]. Teaching and learning has become more manageable and diverse as students can perform many learning activities freely and easily, for instance, they can download lectures notes almost instantaneously for lectures that they had missed. Predictably, mobile learning systems based on android technology are poised to dominate the M-learning realm given the rich, appealing multimedia contents such as audio, videos, animations that can be downloaded effortlessly into students' mobile devices.

IV. LEARNING ENVIRONMENTS

Mobile learning is a form of digital learning which can be applied for teaching and learning purposes where some educational experts view it as a subset of e-learning but with a subtle difference—contents are delivered onto mobile devices rather than the ubiquitous desktop personal computers. Teaching and learning by using android platform can be easily implemented without heavy computing investment. There are several factors that make mobile computing as an appealing platform. First, android operating system to run the mobile devices is conveniently and freely available, thus making installation a simple, neat process.

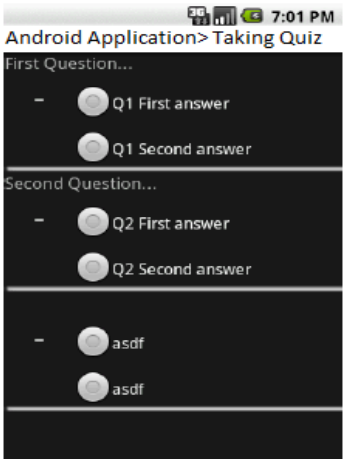


Figure 2. Example of Interface for Quiz

Second, there is a huge application base of learning materials and contents, which is continually expanding, that can be easily accessed by students and instructors alike. For example, students can download and practice short tests or quizzes on their mobile phones where prompt feedback is instantly displayed to improve comprehension This type of learning occurring in short bursts is appealing to young generation [7]. Figure 2 depicts such a test available on students' mobile phones that asks simple questions pertaining to a particular information technology course. In addition, students can download notes from Google doc website using android platform. Currently, the technology enables students to share and edit documents online collaboratively; thus, the notion of collective intelligence has transformed from an abstract concept into tangible realization in the educational realm. Figure 3 shown below illustrates the interface for students to download documents over the related website.

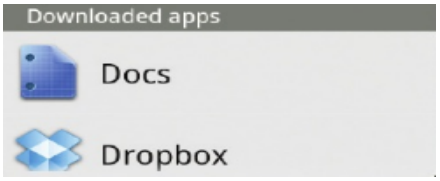


Figure 3. Google Doc

Third, there exists a repository of knowledge for sharing information among practitioners of mobile learning that contributes to the expansion of M-learning in campuses. This creates a community of practice that continually enhances the technical capabilities of M-learning systems irrespective of their background.

In spite of the many benefits accorded by M-learning, there are still some issues that need to be addressed by many concerned. For example, it is not that easy to measure the actual level of meaningful learning that takes place using this type of learning at best, and such a mode of learning can be easily abused at worst. Assimilating this technology into normal classroom activities requires not only structural changes but new thinking regarding learning of this nature is urgently required [8]. This problem is further compounded as individuals involved in M-learning come with different knowledge, skills, and expertise; and understandably, their

involvements in this type of learning will be quite unpredictable to judge [9].

V. RESEARCH METHODOLOGY

This study was conducted by means of a survey to measure undergraduates' perception of M-learning used throughout a semester long, beginning May 2011 until September 2011. Acknowledgement about the course contents and courses program was carried out by text messaging or SMS. The survey was conducted involving a total of 56 students were voluntarily participated in this study. They were divided into two different groups: the first group comprised undergraduates who used the university' e-learning; and the second group consisting of undergraduates who used their mobile phones, running on android 2.3.2, android 2.3.3 and android 2.3.4, to gain access to learning contents.

For the survey, the questionnaires were prepared in Google doc form, and the elicited data based on respondents' responses were collected from Google doc online database. For the learning materials, the two groups of students had to download notes from two different platforms. The first group downloaded learning materials from the e-learning system of the university, whilst the second group of students downloaded similar materials from Google Doc by using their mobile phones.

A. Instrument of Research

The questionnaires comprised 4 sections pertaining to respondents' demographic, android learning environment system, e-Learning environment provided by the university, and the effectiveness of the android learning system. For the respondents' demographic, the data collected were gender, age, and academic achievement. For the Android learning environment, data gathered were related to respondents' perceptions on its features and usability. Likewise, the same perceptions of the respondents were elicited for the e-Learning environment system.

Respondents were asked for their opinions on the items of the questionnaire based on Likert-type scales as follows: 1 for strong disagreement, 2 for disagreement, 3 for being neutral, 4 for agreement, and 5 for strong agreement. All data collected were analyzed by using Statistical Package for Social Science (SPSS) version 12. Descriptive statistics for the demographic were based on frequency; for items pertaining to systems' features, mean scores were calculated to measure the respondents' responses.

B. Findings and Discussion

Table I as shown below summarizes the descriptive statistics for the respondents demographic based on frequency counts.

Table II summarizes all the respondents' responses pertaining to their perceptions on the features provided by the novel learning systems, namely the android mobile learning system and the university's e-learning learning system.

All the five items of the questionnaires were analyzed to reveal mean scores based on the 56 respondents' responses. For items 1 and 2 (see Table II), the mean scores for the university's e-learning system and android mobile learning system were 4.0 and 4.5, respectively. Clearly, the respondents that had used the mobile learning system were more receptive to using the system where they regarded the system to be easy to use and also to be interesting.

TABLE I. DEMOGRAPHY OF THE STUDENTS INCLUDING THE ANDROID PLATFORM

Item	Range	Frequency
Gender	Male Female	20 36
Age (in years)	18-22 22-26	10 46
Respondents' academic achievements	CGPA < 3 CGPA > 3	15 41
Versions of Android Operating Systems	Gingerbread 2.3.2 Gingerbread 2.3.3 Gingerbread 2.3.4	10 20 26

TABLE II. RESPONDENTS' PERCEPTIONS ON THE LEARNING SYSTEMS' FEATURES

Statements	Mean
The university's e-learning environments is interesting and fun to use	4
Android's mobile learning environment is interesting and fun to use	4.5
Android's mobile learning environment provides more space for self-learning	4.5
I can easily download lecture notes from Google Doc by using my mobile phone at any time and place	5
I can easily download lecture notes from the university' e-learning system at any time and place	3.5

More poignant, all respondents that used the mobile platform were unanimous that they could use their hand phones for self-learning where item 3 of the questionnaire recorded a mean score of 4.5. This finding is not surprising as android-based mobile phones are quite affordable to own, and in terms of performance, they are very stable and could perform all the necessary functions with greater ease. These two factors seem to be a driving force to spur greater growth of mobile learning in the future.

For items 4 and 5 (see Table II), in terms of having the capability to download lecture materials, the mobile learning group and the e-learning group recorded mean scores of 5 and 3.5, respectively. Evidently, this feature of greater capability for downloading is better accorded by mobile learning platform compared to e-learning platform. Apparently, there are several reasons why mobile learning is greatly favored for this feature of a digital learning system. First, a learning system that guarantees uninterrupted access to learning materials can ensure smooth flow of learning process where students can download any documents at any time, no matter where they are. This partly contributes to a more conducive learning environment that suits the needs of today's younger generation: the digital natives.

On the other hand, the availability of lecture materials by many e-learning systems is sometimes compromised by technical problems with most cases resulting in breakdowns, which hinder constant access for online documents. Thus, the

feature of having full access of learning materials at all time will be a decisive factor that favors a mobile learning system over an e-learning system when the target groups of learning are those of adolescent age.

Another factor that seemed to motivate students to use the mobile learning system as compared to the university's e-learning system rested on the fact that the learning environment based on android itself was more interactive and simple to interact with. These students had ample time to download lecture notes from the Google Doc website without interruptions, and they could take quizzes and short tests at leisure, which further enriches their learning experience. Thus, learning becomes more fun. This is not unexpected as there is a vast repository of learning tools, widgets, and applications that could be accessed freely or bought at minimal cost from the android market. Once downloaded, all these digital materials can be utilized instantly and repetitively. Moreover, those students that have these materials can share with their peers by exchanging files through the Bluetooth technology, which is one of the standard features of today's mobile devices.

In sum, this research suggests that mobile learning can be quite easily implemented as clearly demonstrated by the respondents involved in this study. Mobile learning can be cost-effectively implemented as android operating system used to run the mobile phones is freely available. Moreover, newer, better versions of this operating system are constantly update, giving better performance in terms of processing and intuitive interface design. Mobile learning systems powered by android technology can make learning more fun, interactive and intuitive. This mobile learning system can be used by educational practitioners, such as instructors and teachers, to prepare the study notes in any standard digital formats and then upload them onto the Google Doc website, where they can be accessed and shared. Self-paced learning and collaborative learning can be realized with ease to improve the learning process that befits today's challenging learning environments.

VI. Future Research

Future research is needed to examine the full impact of mobile learning both from the technological and pedagogical perspectives. Expectedly, the introduction of any new, novel technology would have profound impact, affecting both students and educators. Students will be overwhelmed with the technological gizmo that is normally designed for routine chores, not for educational purposes. Thus, proper working ethics and code of practices are entailed to ensure optimal use of mobile devices for mobile learning. Likewise, educators must keep abreast with latest technologies to make efficient use of them. Therefore, future research should focus some of these issues to help realize digital learning environments that complement the conventional learning approach.

VII. Conclusion

The use of mobile technologies, in particular the non-proprietary android technology, offers many educational opportunities to the stake holders: the students, the instructors, and the administrators. However, as for today, there are many emerging information and communication technologies entering the educational realm that forces practitioners to rethink how this novelty can be judiciously applied to improve the overall learning process. Many educational benefits of the novelty can be easily identified; however, realizing these is not a straightforward process as there is a web of interrelated factors that needs delicate unweaving to ensure effective and efficient implementations in educational institutions.

References

[1] Keegan, D. (2002). The future of learning: From e-learning to m-learning.
Retrieved September 7th, 2002 from the World Wide Web: http://learning.ericsson.net/leonardo/thebook/chapter4.html#milearn

[2] Barbosa and Geyer (2005). Pervasive personal pedagogical agent: A mobile agent shall always be a learner. Proceeding IADIS International Conference Mobile Learning, Malta 281-285 Bobnano

[3] Brown, H.T(2005),"Towards a model for MLearning", International Journal on E-Learning, 4(3),299-315

[4] Traxler, J. (2005). Institutional issues: Embedding and supporting. In A. Kukulska-Hulme & J.Traxler (Eds.), Mobile learning: A handbook for educators and trainers (pp. 173-188), London:Routledge.

[5] Shanmugapriya M.& Tamilarasia, A.(2011), Designing an m-learning application for ubiquitous learning environment in the android based mobile devices using web services, Indian Journal of Computer Science and Engineering(IJCSE) ,22-30

[6] Williams, A.J &Pence, H.E (2011). Smart phones, a powerful tool in Chemistry classroom, Journal of Chemical Education, 88, 683-686.

[7] P. Pocatilu, F. Alecu and M. Vetrici, Measuring the Efficiency of Cloud Computing for E-learning Systems, WSEAS TRANSACTIONS on COMPUTERS, Issue 1, Volume 9, January 2010, pp. 42-51.

[8] Winters, N.& Mor.Y (2008)'IDR: a participatory methodology for interdisciplinary design in technology enhanced learning' (Computers & Education, 50(2),579-600))

[9] F. Alecu, P. Pocatilu and S. Capisizu, WiMAX Security Issues in E-Learning Systems, Proc. of 2nd International Conference on Security for IT & C in Journal of Information Technology and Communication Security, Bucharest, November 2009, pp. 45-52

Authors Profile

Hafizul Fahri Hanafi is a lecturer in the Computer Department of Sultan Idris Education University, Malaysia. He specializes in Software Engineering and E-Learning Technologies
(Email: apiltzs@gmail.com)

Khairulanuar Samsuddin is a lecturer in the Computer Department of Sultan Idris Education University, Malaysia. He specializes in Virtual Reality in Education and Engineering
(Email: khairul@fskik.upsi.edu.my)

Simple and Efficient Contract Signing Protocol

Abdullah M. Alaraj
Information Technology Department
College of Computer, Qassim University
Saudi Arabia

***Abstract*—In this paper, a new contract signing protocol is proposed based on the RSA signature scheme. The protocol will allow two parties to sign the same contract and then exchange their digital signatures. The protocol ensures fairness in that it offers parties greater security: either both parties receive each other's signatures or neither does. The protocol is based on offline Trusted Third Party (TTP) that will be brought into play only if one party fails to sign the contract. Otherwise, the TTP remains inactive. The protocol consists of only three messages that are exchanged between the two parties.**

Keywords-contract signing; fair exchange protocol; digital signature; protocols; security.

I. INTRODUCTION

Contracts play an important role in many business transactions. Traditionally, paper-based contracts are signed by the transacting parties who need to be present at the same venue and at the same time. Each party signs a copy of the contract for every contracting party so that every party has a copy of the signed contract.

If the parties, however, are not able to meet to sign the paper-based contract, then signing an electronic contract is an alternative. The problem with signing electronic contracts, however, is exchanging the signatures of the parties, especially where there is a lack of trust between parties. One party may send the other party their signature on the contract but may not receive the signature of the other party in return. To solve the problems of exchanging digital signatures, contract signing protocols are used [3, 4, 5, 9, 10]. Contract Signing Protocols ensure that either contracting parties receive each other's signature or none does.

In this paper, a new, efficient contract signing protocol is proposed. The proposed protocol is based on offline trusted third party (TTP) that brought into play only if one party fails to send their signature on the contract. In the normal execution of the protocol, the two parties will exchange their signatures directly.

This paper is organized as follows. Related work is presented in section II. Section III presents the proposed protocol that comprises the exchange protocol and dispute resolution protocol. The analysis of the proposed protocol is discussed in section IV. The comparison of the proposed protocol with related protocols is presented in section V.

II. RELATED WORK

Early contract signing protocols (as in [7, 16]) allow the parties to exchange their signatures directly without any involvement from third party. That is, the parties gradually exchange their signatures in part until both signatures are complete. If one party fails to send an additional part of the signature, the other party works to search for that remaining part. The gradual exchange protocols are based on the assumption that the two parties have the same computational power to ensure fairness. However, in most applications this assumption is not realistic [5]. The gradual exchange protocols require a large number of rounds to complete the exchange of signatures.

To overcome the problems of gradual exchange of signatures, a trusted third party (TTP) is used in contract signing protocols. The TTP helps the contracting parties to exchange their signatures in a reliable and secure manner. The TTP can be used online or offline.

In the online-based third party contract signing protocols [as in 6, 8,10] the TTP will be actively involved in the exchange of the signatures between the parties. The parties will sign the contract and send their signatures to the TTP who will verify the signatures and if they are correctly verified the TTP will forward the signatures to the parties. The main problem with this approach is that the TTP is involved in every exchange and this may create a bottleneck. In addition to this, the fees of the third party make this a costly approach.

In the offline-based third party contract signing protocols [as in 3, 4, 5, 11, 13 (also called optimistic – 11)], the parties will directly exchange each other's signatures on a contract. If one party fails to submit their signature, the third party will be brought in to resolve any dispute. In the offline-based third party contract signing protocols, the TTP is rarely involved which reduces the cost of running TTP. Also, the turnaround time is eliminated since the parties exchange their signatures directly.

A category of offline TTP-based contract signing protocols has been proposed [3, 4, 5]. This category overcomes the farness problem by using verifiable and recoverable encrypted signatures. This approach will generally work as described below. Let's say that two contracting parties, Alice and Bob, want to exchange their signatures on a contract.

Alice will sign the contract, encrypt the signature and then send the encrypted signature to Bob. Bob will then verify the encrypted signature and if it is correctly verified, send his signature to Alice. If Alice finds that Bob's signature is correct then she will send the decryption key to Bob to decrypt her encrypted signature. If Alice fails to send the decryption key, Bob will contact the TTP to recover the decryption key.

Nenadic, Zhang and Barton[3] proposed a fair signature exchange protocol. The protocol is based on the verifiable and recoverable encryption of signatures on a contract. Alice will send her partially encrypted signature to Bob who will be able to verify it. If the encrypted signature is correctly verified then Bob will send Alice his signature. On receiving Bob's signature, Alice will verify it and if it is correctly verified then Alice will send the decryption key to Bob to decrypt the encrypted signature. If Alice does not send the decryption key, Bob will contact the TTP to recover Alice's signature.

Ateniese [4] also proposed a fair contract signing protocol. Ateniese's protocol is based on the verifiable and recoverable encryption of a signature. If Alice and Bob want to exchange their signatures on a contract then the protocol will work as follows. Alice will first sign the contract, then encrypt the signed contract with the public key of the trusted third party (TTP). Alice will then send Bob: (1) the encrypted signature, (2) evidence stating that Alice has correctly encrypted her signature on the contract. On receiving Alice's message, Bob will verify the evidence. If the evidence is valid then Bob will send his signature on the contract to Alice. On receiving Bob's signature, Alice will verify it and if it is valid then Alice will send her signature on the contract to Bob. If Alice does not send her signature to Bob or Alice's signature is invalid then Bob can contact the TTP to resolve the dispute.

Wang [5] proposed a protocol for signing contracts online. Their protocol is based on the RSA signature. If Alice and Bob are planning to exchange their signatures on a contract using Wang's protocol [5] then Alice will first split her private key into two parts d1 and d2. Only d2 will be sent to TTP. Alice will send Bob her partial signature that was signed using d1. On receiving Alice's partial signature, Bob will initiate an interactive zero-knowledge protocol with Alice to check whether Alice's partial signature is correct. If it is correctly verified then Bob will send his signature to Alice. After Alice receives Bob's signature, Alice will verify it and if it is correctly verified then Alice will send Bob the second part of her signature. If, however, Alice did not send the second part of the signature, Bob can contact the TTP to resolve the dispute.

In this paper, we propose a new approach that uses verifiable and recoverable encryption of signatures that will allow the party who receives the encrypted signature to verify it. If he / she correctly verifies the encrypted signature, then it is safe for this party to release his / her signature to the other party because the TTP can be contacted to recover the signature if the other party fails to submit his / her signature. The proposed protocol does not use the interactive zero-knowledge proofs for verifying the encrypted signature as in [4 & 5]. Rather, the contract certificate that is introduced in this paper will allow the party who receives the encrypted signature to verify it.

III. The Proposed Contract Signing Protocol

A. Notations

The following represents the notations used in the proposed protocol:

- P_a, P_b, and P_t: parties a, b, and TTP, respectively.
- C: The contract to be signed by P_a and P_b
- $C._{at}$: the certificate for the shared public key between P_a and P_t. $C._{at}$ is issued by P_t. A standard X.509 certificate [12] can be used to implement $C._{at}$
- $Pk_x = (e_x, n_x)$: RSA Public Key [14] of the party x, where n_x is a public RSA modulus and e_x is a public exponent
- $Sk_x = (d_x, n_x)$: RSA Private Key [14] of the party x, where n_x is a public RSA modulus and d_x is a private exponent
- h(M): a strong-collision-resistant one-way hash function
- $enc.pk_x(M)$: an RSA [14] encryption of message M using the public key pk_x (e_x, n_x). The encryption of M is computed as follows: $enc.pk_x(M) = M^{ex} \bmod n_x$
- $enc.sk_x(Z)$: an RSA [14] decryption of Z using the private key sk_x (d_x, n_x). The decryption of Z is computed as follows: $enc.sk_x(Z) = Z^{dx} \bmod n_x$
- $Sig._x(M)$: the RSA digital signature [14] of the party *x* on M. The digital signature of party *x* on M is computed by encrypting the hash value of M using the private key $sk_x(d_x, n_x)$.
- C-Cert: the contract certificate. C-Cert is issued by CA. The contents of C-Cert are:
    - heSig: the hash value of the signature of P_a on the contract encrypted with pk_{at} i.e. "$h(enc.pk_{at}(Sig._a(C)))$"
    - hC: hash value of the contract
    - CA's signature on C-Cert
- $P_x \rightarrow P_y$: M, means party x sends message M to party y
- X + Y: concatenation of X and Y

B. Assumptions

The following represents the assumptions used in the proposed protocol:

- Channels between P_a, P_b and P_t are resilient i.e. all sent messages will be received by their intended recipients
- Parties will use the same hashing, encryption, decryption algorithms.
- P_t is trusted by all parties and will not collude with any other party
- Parties P_a and P_b will agree on the contract before the protocol starts

- Parties (P_a, P_b and P_t) already have their public keys and they are certified from CA

C. Registration

In the registration phase, Pa needs to do the following:

- P_a will request from P_t to share an RSA public key with it. The shared public key is denoted as $pk_{at} = (e_{at}, n_{at})$ and its corresponding private key is denoted as $sk_{at} = (d_{at}, n_{at})$. P_t will certify the shared public key and issue the shared public key certificate $C._{at}$
- Pa will sign the contract "C" using its private key sk_a as $Sig._a(C)$ and then send the following to CA to certify the encrypted signature and issue C-Cert:

$Sig._a(C) + C + C._{at}$

On receiving P_a's request, CA will verify if the received signature is for the contract C included in P_a's message. If so, then CA will encrypt $Sig._a(C)$ using the shared public key pk_{at} that is included in C.at. That is, CA will compute:

$enc.pk_{at}(Sig._a(C))$

Then, CA will issue C-Cert that includes the items mentioned in the "Notations" section.

D. Exchange Protocol

The exchange protocol represents the normal execution of the protocol. It consists of the following three steps (see Fig. 1):

1. [E-M1]: $P_a \rightarrow P_b$: $C + C._{at}$ + C-Cert + $enc.pk_{at}(Sig._a(C))$
2. [E-M2]: $P_b \rightarrow P_a$: $Sig._b(C)$
3. [E-M3]: $P_a \rightarrow P_b$: $Sig._a(C)$

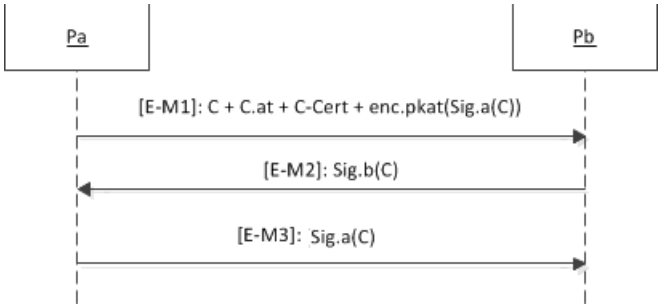


Figure 1. Exchange Protocol

Step [E-M1]: P_a encrypts the signed contract with the shared public key pkat. Pa then sends the items C, C.at, C-Cert, $enc.pkat(Sig._a(C))$ to P_b.

Step [E-M2]: once P_b receives E-M1 then they will do the following verifications:

1. P_b will verify the correctness of both $C._{at}$ and C-Cert by verifying the signatures on these certificates.
2. If the certificates are correctly verified then P_b will compute the hash value of the contract and then compare it with "hC" that is included in C-Cert.
3. P_b will also need to verify the correctness of the encrypted signature of P_a on the contract i.e. P_b will verify "$enc.pk_{at}(Sig._a(C))$". To verify the encrypted signature, P_b will compute the hash value of "$enc.pk_{at}(Sig._a(C))$" then compare it with "heSig" that is included in C-Cert. If they match, it means that P_a encrypted the correct signature.

If all verifications are correct then P_b will sign the contract using their private key skb then will send the signed contract "$Sig._b(C)$" to P_a.

Step [E-M3]: once P_a receives Sig.b (C), Pa will verify P_b's signature. That is, P_a will decrypt the signature to get the hash value of the contract then compare it with "hC" that is included in C-Cert. If P_b's signature is correctly verified then P_a will send their signature Sig.a(C) to P_b

Once P_b receives $Sig._a(C)$ then P_b will verify it by decrypting the signature to get the hash value of the contract and compare it with "hC" that is included in C-Cert. If the verification is correct then the received signature is correct.

Now, both P_a and P_b have each other's signatures on the contract. Therefore, fairness is ensured. If P_a did not send E-M3 or sent incorrect E-M3 then P_b can contact P_t using the dispute resolution protocol to resolve the dispute.

E. Dispute Resolution Protocol

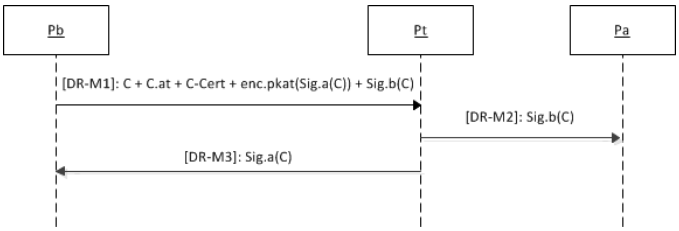


Figure 2. Dispute Resolution Protocol

If P_b did not receive the step E-M3 or received an incorrect E-M3, P_b can contact P_t to resolve the dispute. The dispute resolution protocol consists of the following three steps (see Fig. 2):

1. [DR-M1]: $P_b \rightarrow P_t$: $C + C._{at}$ + C-Cert + $enc.pk_{at}(Sig._a(C)) + Sig._b(C)$
2. [DR-M2]: $P_t \rightarrow P_a$: $Sig._b(C)$
3. [DR-M3]: $P_t \rightarrow P_b$: $Sig._a(C)$

Step [DR-M1]: if P_b did not receive the correct signature or did not receive the signature at all then P_b will send message DR-M1 to P_t to request a resolution.

Step [DR-M2]: once P_t receives DR-M1 then they will do the following verifications:

- P_t will verify the correctness of $C._{at}$ and C-Cert by checking the signatures on these certificates.
- If the certificates are correctly verified then P_t will verify the correctness of the encrypted signature of P_a on the contract i.e. $enc.pk_{at}(Sig._a(C))$. To verify the encrypted signature, P_t will either (i) compute the hash value of $enc.pk_{at}(Sig._a(C))$ then compare it with

"heSig" that is included in C-Cert. If they match it means that P_a encrypted the correct signature, or (ii) P_t has the private key "sk_{at}" corresponding to the shared public key so it can decrypt the encrypted signature i.e. $enc.pk_{at}(Sig._a(C))$ and then decrypt the signature with pk_a and compare the decrypted hash with "hC" that is included in C-Cert.

- P_t will also verify $Sig._b(C)$ by decrypting the signature with pk_b then comparing the decrypted hash with "hC" that is included in C-Cert.

If all verifications are correct then P_t will send the message DR-M2 to P_a and DR-M3 to P_b. DR-M2 includes the signature of P_b on the contract.

The signature of P_b on the contract is sent to P_a to ensure fairness in the case where P_b contacted P_t after receiving E-M1 i.e. P_b may cheat by contacting P_t before sending E-M2 to P_a.

Step [DR-M3]: P_t will send $Sig._a(C)$ to P_b in DR-M3

Now, both P_a and P_b have each other's signature on the contract. Fairness is ensured either in the exchange protocol or in the dispute resolution protocol if P_a acts dishonestly.

IV. Analysis

The fairness property in our protocol will be evaluated by studying the following four cases: (1) the first case where P_a is honest and P_b is dishonest, (2) the second case where P_a is dishonest and P_b is honest, (3) the third case where both P_a and P_b are dishonest, and (4) the forth case where both P_a and P_b are honest.

- Case 1: If P_a is honest and P_b is dishonest. P_b acts dishonestly by sending an incorrect signature to P_a or by contacting P_t before sending his signature to P_a. In the first scenario where P_b sends an incorrect signature to P_a, P_a will check P_b's signature. Then if it is incorrect, P_a will not send his signature to P_b in E-M3. In the second scenario where P_b contacted P_t before sending his signature to P_a, P_t will check P_b's request and if it is correctly verified then P_t will send the resolution to both P_a and P_b. Therefore, fairness is ensured
- Case 2: P_a is dishonest and P_b is honest. P_a can act dishonestly by sending the incorrect E-M1, sending the incorrect E-M3 or not sending the E-M3 at all. In the scenario where P_a sends incorrect E-M1, P_b will verify E-M1 as described in section III. If P_b finds that E-M1 is incorrect, they will not send their signature to P_a in E-M2. In this scenario no one reveals their signature at this stage. In the scenarios where P_a sends incorrect E-M3 to P_b or P_a does not send E-M3, P_b can contact P_t to recover P_a's signature.
- Case 3: both P_a and P_b are dishonest. P_a can act dishonestly by sending the incorrect E-M1, sending the E-M3 or not sending the E-M3 at all. P_b can act dishonestly by sending an incorrect signature to P_a or by contacting P_t before sending his signature to P_a. The scenarios of case 3 are discussed in cases 1 and 2 above.
- Case 4: both P_a and P_b are honest. If both P_a and P_b act honestly then fairness will be ensured in the exchange protocol and there is no need to contact P_t at all.

Therefore, the above analysis of the four cases shows that the fairness is ensured either in the exchange protocol or in the dispute resolution protocol.

It is worth mentioning that P_t does not need to receive any message from P_a in order to resolve any dispute raised by P_b. Rather, P_t will receive the dispute request from P_b and then will decide if P_b's request is valid or not. If the request is valid then P_t will send the resolution electronically to both P_b and P_a.

The certificate C-Cert is unique for each exchange. That is, every time P_a and P_b need to exchange their signatures on a contract then a new certificate will be used. The shared public key certificate $C._{at}$, however, can be used for signing an unlimited number of contracts.

P_t is passive during the exchange protocol i.e. in the normal execution of the protocol P_a and P_b will not need to contact P_t. In case P_a misbehaves then P_t will be contacted by P_b to resolve the dispute.

V. Comparison With Related Work

The proposed protocol will be compared against contract signing protocols that are based on verifiable and recoverable encryption of signatures, namely, Nenadic, Zhang and Barton protocol [3], Ateniese's protocol [4] and Wang's protocol [5].

For the comparison, we analyze the number of messages and the number of modular exponentiations in both the exchange protocol and dispute resolution protocol. The exponentiation is the most expensive cryptographic operation in the finite field [5].

Both the proposed protocol and Ateniese's Protocol [4] have three messages in the exchange protocol whereas Wang Protocol [5] has seven messages. All protocols have three messages in the dispute resolution protocol.

Regarding the modular exponentiations in the exchange protocol, the proposed protocol has the lowest number of modular exponentiations, with only six. Nenadic, Zhang and Barton protocol [3] has the lowest number of modular exponentiations in the dispute resolution protocol with only five modular exponentiations. Our protocol has seven modular exponentiations in the dispute resolution protocol.

Ateniese's Protocol [4] and Wang's protocol [5] require interactive zero-knowledge proofs to allow one party to verify the encrypted signature of the other party. Our protocol offers greater efficiency in that it allows the receiving party to verify the encrypted signature using the contract certificate (C-Cert).

From Table 1, it is clear that the proposed protocol is more efficient compared with the related protocols except for the dispute resolution protocol as Nenadic, Zhang and Barton [3] protocol has the lowest number of modular exponentiations.

TABLE I. PROTOCOLS COMPARISONS

	Nenadic protocol [3]	Ateniese Protocol [4]	Wang Protocol [5]	Our protocol
# messages in exchange protocol	4	3	7	3
# messages in dispute resolution protocol	3	3	3	3
# modular exponentiations in exchange protocol	19 (taken from [3])	22 (taken from [3])	10.5 (taken from [5])	6
# modular exponentiations in dispute resolution protocol	5 (taken from [3])	≥ 20 (taken from [3])	Not mentioned	7

VI. CONCLUSION

A new offline TTP-based fair contract signing protocol is proposed in this paper. The proposed protocol ensures the exchange of signatures of two parties on a contract. At the end of the execution of the protocol, both parties get each other's signatures or neither does. The proposed protocol comprises of only three messages in the exchange protocol as well as only three messages in the dispute resolution protocol. If one party evades during the execution of the protocol, the protocol provides an online resolution for the disputes where the TTP will be involved. The proposed protocol is efficient as it has the lowest number of modular exponentiations in the exchange protocol. In a future study, we plan to investigate how to make the protocol an abuse-free protocol as Wang did in [5]. We also intend to implement and integrate the proposed protocol with e-commerce applications for the exchange of digital signatures between two parties.

REFERENCES

[1] A. Alaraj, "Optimizing One Fair Document Exchange Protocol" International Journal of Network Security & Its Applications (IJNSA), Vol.4, No.1, pp. 1- 12, January 2012

[2] A. Alaraj and M. Munro, "An e-Commerce Fair Exchange Protocol that Enforces the Customerto be Honest". International Journal of Product Lifecycle Management, IJPLM, Vol.3, Nos.2/3, pp. 114-131, 2008

[3] A. Nenadic, N. Zhang, and S. K. Barton, "A Secure and Fair DSA-based Signature Exchange Protocol ", the 9th IEEE Symposium on Computers and Communications (ISCC'2004), Alexandria, Egypt June 29-July 1, 2004, pp. 412-417.

[4] G. Ateniese, "Efficient verifiable encryption (and fair exchange) of digital signature," in Proc. ACMConf. Computer and Communications Security (CCS'99), 1999, pp. 138–146, ACM Press

[5] G. Wang"An Abuse-Free Fair Contract-Signing Protocol Based on the RSA Signature" by G. Wang, IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 5, NO. 1, pp. 158-168, MARCH 2010

[6] H. Burk and A. Pfitzmann, "Value Exchange Systems Enabling Security and Unobservability", Computers & Security 9, pp. 715-721, 1990

[7] I. Damgard, "Practical and provably secure release of a secret and exchange of signatures". In: Proceedings of advances in cryptology – EUROCRYE'T '93, vol. 765. Berlin, Germany: LNCS, Springer-Verlag; 1994. pp. 200–17

[8] J. Zhou and D. Gollmann, "A fair non-repudiation protocol," in Proc. IEEE Symp. Security Privacy, 1996, pp. 55–61, IEEE Computer Press

[9] L. Harn and C. Lin "Contract signature in e-commerce" Computers and Electrical Engineering37 (2011), pp. 169-173, 2011

[10] M. Ben-Or, O. Goldreich, S. Micali, and R. Rivest, "A Fair Protocol for Signing Contracts",IEEE Transactions on Information Theory, vol. 36, no. 1, pp. 40-46, Jan. 1990

[11] N. Asokan, M. Schunter, and M. Waidner, "Optimistic Protocols for Fair Exchange", Proc.Fourth ACM Conf. Computer and Communication Security, pp. 8-17, Zurich, Switzerland,April 1997

[12] Public-Key Infrastructure (X.509), The PKIX working group, available athttp://datatracker.ietf.org/wg/pkix/charter/ accessed on 16-02-12

[13] Q. Shi, N. Zhang, M. Merabtia: Fair exchange of valuable information: A generalised framework. Journal of Computer and System Sciences 77 (2011),pp.348–371

[14] R. Rivest, A. Shamir, L. Adleman "A method for obtaining digital signatures and public-keycryptosystems", Commun ACM 1978; pp. 120–126, 1978

[15] S. Micali, "Simple and fast optimistic protocols for fair electronic exchange," in Proc. PODC'03, 2003, pp. 12–19, ACM Press.

[16] T. Okamoto and K. Ohta. "How to simultaneously exchange secrets by general assumptions". In: Proceedings of ACM conference on computer and communication security, 1994. pp. 184–92

[17] X. Liang, Z Cao, R. Lu, and L Qin "Efficient and secure protocol in fair document exchange", Computer Standards & Interfaces, Vol. 30 (2008), pp. 167–176, 2008

[18] Z. Shao "Security analysis of two RSA-Based fair document exchange protocol". In: Proceedings of the Second International Workshop on Computer Science and Engineering, Qingdao, China, pp. 55-59, 2009

AUTHOR PROFILE

Abdullah Alaraj is presently a faculty member in the department of Information Technology, College of Computer, Qassim University, Saudi Arabia. He received his BSc in Computer Science from King Saud University (Saudi Arabia), his MSc in Internet and Distributed Systems from Durham University(UK), and his PhD from Durham University (UK). His areas of research interests include: e-commerce security, fair exchange protocols, fraud, trust, information security

The Use of Information and Communication Technologies (ICT) in Front Office Operations of Chain Hotels in Ghana

[*1]Albert Kwansah Ansah, [2]Victoria S. Blankson
[1]Department of Computer Science and Engineering,
University of Mines and Technology (UMaT), Tarkwa,
Ghana

[1]Millicent Kontoh
[2]Department of Consideration of Tourism and Hospitality,
Zenith University College, Accra, Ghana

***Abstract*—The proliferation of Information and Communication Technologies (ICT) coupled with sophisticated network protocols have unveiled new avenues for enterprises and organizations and the hospitality industries cannot be left out. Technology-based systems stand in a pivotal position to offer better service to the populace. Hospitality industries such as hotels can take advantage of the pervasiveness of ICT vis-à-vis technology-based systems to advance some of their operations. This paper seeks to assess the use of Information and Communication Technologies (ICT) in a front office operation of chain hotels in Ghana. The paper determines the extent of the use of information technology in a front office operation of chain hotels in Ghana. The paper continues to assess the effect of the use of information technology in the front office operation of chain hotels in Ghana, thus if the use of ICT has any effect on chain hotels' front office operations. The paper further makes recommendations to chain hotel operators and the Ghana Tourist Authority (GTA) and policy makers on the use of information and communications technology in front office operation in chain hotels. Three chain hotels in Ghana were assessed.**

Keywords-Front Office Operation; ICT; Chain Hotels; Electronic Point of Sale; Reservation.

I. INTRODUCTION

Information technology is everywhere in this day and age and adding to communication technology, the possibility of how it can be used is endless. Information and Communications Technology (ICT) involves the use of computer hardware, software and telecommunication devices to store, manipulate, convert, protect, send and receive data (Olifer and Olifer, 2006). Computer and technological devices have made it easier for professionals to collect, store, manipulate and share data and information both individually and within organizations, small and large, public and private. ICT has become an integral part of human daily activities that sometimes we barely notice its effect, and has had a major impact on the way we live, work and play. The way hotel companies sell to consumers is changing dramatically over the past few years. The use of networking in front office helps very much in sharing information across the hotels in chain. The growth in the use of networked computers is one of the most significant trends in modern computing. Though interconnection of computers in itself is not new, application of networking is seeing a dramatic increase such that it is now one of the major issues in computer and information technology (Ritchie, 2003). Computers are presently widely used in sectors such as banking, education, health, commerce, agriculture, transport, communication. Computers in business operations could be used for typing documents, keeping and retrieving information, data analyses, sending and receiving information over short and long distances. Computers are becoming popular in businesses because work done by them could be very fast, thus results in increasing productivity, accurate in performing repetitive task, store information safely in relatively small space (Olifer and Olifer, 2006).

Despite the increasing popularity of computers, they are still limited to those with the skill to use them. In hotels, computers are used in Accounting for guest, data management, forecasting guest demand for reservations, management of guest services, revenue and reservation management and yield management. These functions could be performed with the aid of management software (Abbott and Lewry, 1999).

The delivery of businesses today is exposed to information and communication technologies either directly or indirectly. Hotels being a subset of the hospitality industry traditionally practise manual system of operation in their front offices as from occupancy of the guest rooms, recording of guest expenditure through to the eventual departure of the guest. Today, these operations are done using the electronic system. The use of information technology in a front office operation of the hospitality industry is fast growing, thus making work easy. Computerization is becoming very important for the efficient and effective operations of the front office.

II. LITERATURE REVIEW

A. ICT in Front Office Operations

Technology is used to push slower moving businesses by providing better service, improved decision making and increasing revenue. Hotel operators are now realising that a brand in itself is not enough (Knowles, 1998). Networks are used to link together computers, storage devices, printers, telephony and other electronic devices (Odom, 2007). The hotel's systems could be networked to share resources via either Local Area Network (LAN) or Wide Area Network (WAN). The LAN helps to share information within a hotel, such as from restaurant to front office and the WAN helps to

share information from one hotel branch to another within the same chain across different geographical areas. Computers can connect to these networks to use facilities from another hotel or location.

Through ICT, Centralised Reservation Systems (CRS) could be used to exploit data and information resources. The link to a centralised reservation system is considered one of the most important benefits of joining any hotel franchise (Knowles, 1998). Networking the centralised reservation system enhances cost effectiveness, faster communications, and effective exchange of information and efficient management of data (Lucey, 2005).

With a sophisticated central reservation system, a hotel chain provides individual hotels and managers in the chain with a tool to increase reservations, maximize sales, implement yield management, enhance market capabilities and improve quest services. The systems are also linked to airline CRSs to form a Global Reservation System in order to allow travel agents to make direct reservations for their clients. CRS in no doubt faces unprecedented operational and guest services challenges such as system downtime, but may still be essential to survival (Knowles, 1998).

With the numerous positive impact of ICT in businesses, it is hard to imagine a contemporary business functioning without adoption of ICT. ICT permeates every aspect of twenty first century businesses. In hotel front office, computers are used to create bills and invoices, to monitor bookings and reservations, to check-in and check-out guests, to record guest expenditure and share information within and across the hotel. Using IT, guests could stay at the comfort of their homes and private places, communicate with the front office staff via telephony or the Internet to make and confirm reservations. Credit and debit card payments have become an integral part of front office operation today with appropriate software and hardware. Guests can make payment for their booking online to facilitate their reservation, which saves time and reduces queuing at the front offices. Electronic Fund Transfer (EFT) helps quick updates of both hotel and guests' accounts after reservation. Tools such as Computer Managed Learning (CML) and Computer Managed Instruction (CMI) are used as administrative resources to organize guest data, occupancies and vacant rooms.

The way hotel companies sell to customers has changed dramatically over the past few years. In hotel front office operations, the Central Reservation Systems (CRS) is used to share information such as available rooms, room rates etc across hotels within a chain. Global Distribution System (GDS) is also used to link directly the reservation system of hotels, airlines and so on, on a worldwide basis; this can be accessed through seamless connectivity via the internet (Baker et al, 2000).

Technological development certainly would have a great impact on the front office activities. Software packages cover virtually every front office function from reservations, room allocation, and guest history, billing and accounting to the production of management information (Knowles, 1998). It is for this reason that using information and communication technologies in front office operations in chain hotels in Ghana is getting a face-lift with both positive and negative impact. The positive impact may include networking to share information and resources within and across the hotels to enhance check-in of guests and guests' transactions, and easy billing of guests account during check out. With all the good impact of ICT on chain hotel, there should be a budgetary allocation to accommodate the procurement of software, hardware and networking devices and installation, security headaches, training of staff on the use of systems, routine maintenance, redesign of cabling run, on-site systems administrator, disposal of unwanted hardware devices etc. and could bring negative consequence to the hotel owners and managers.

The effect of IT hardware on the staff health cannot be overemphasised. Working with video display terminal (VDT) and the Keyboard can be productive in rewarding and a lot of fun. Unfortunately, prolong postures, coupled with high level of concentration and occasional frustration of things going less than perfect can lead to physical problems like carpal tunnel syndrome (CTS), and computer vision syndrome. Continual clicking and small precise motions involved in mouse use are a repetitive action that could be a health hazard. Improper disposal of unwanted hardware device may also be hazardous to the staff therefore extra money may be spent for apt disposal (Olifer and Olifer, 2006).

The growing importance of computers in the daily lives has raised concerns about possible treat to computers and data. Data collected about clients should be protected from misuse and therefore adequate security measures must be employed, thus data integrity and confidentiality must be ensured. These chain hotels may spend huge amount of money to take care of both hardware and software security measures such as purchasing firewalls and third party backup software to protect data held about their clients or hire backup operators to take charge of data backup backups. Other negative consequence may be losing of huge amount of money during system downtime, that is, when the systems are off-line, clients cannot make reservations both on-line and on the telephone. Chain hotels may also have to spend so much money in training personnel to gain expertise on the use of IT (Olifer and Olifer, 2006). These could affect their budget significantly.

B. Chain Hotel Front Office Operations

The front office department is the most noticeable department in the hotel. It is traditionally known as reception and it is the focal point of most activities within a hospitality business, whether it is a large or small hotel, a cruise liner, a holiday centre, a time-share resort or a youth hostel. The front office is the first and last place where a guest has direct contact with the business, and is the most visible of all departments within the hospitality industry. The front office is a term accepted as including back of house responsibilities, such as switchboard, accounts, cashier and night audit, front desk, concierge and guest services (Edexcel Limited, 2010). The department may have the front desk, reservations, telephony and the concierge, which provides guests with services and facilities. The main function of the front office department is to support and smooth the progress of guest transaction and services through all the four stages in the guest cycle; that is, pre-arrival, arrival, occupancy and departure. The front office

department does all the guest transactions such as reservations, check-in and registration, mail and information, uniformed service and baggage handling, telephone calls and messages, guest accounts, check-out and bill settlement (Baker et al, 2000).

Until the 1990's, nearly all hotels were operating under the manual system. With the introduction of computers, hotels are shifting to automated systems. Most five-star hotels operate under the fully automated system. The fully automated systems are computer-based. This is the best system ever used in the hotel industry and it is characterized by the excessive use of departmental software package programs integrated and connected to a main frame or terminal server situated at the front office department.[1, 2]

There are a wide range of point of sale (POS) applications that are compatible with UNIX and Windows. The availability of processing power, data storage, networking, and graphical user interface made it possible to develop flexible and highly functional POS systems. Some of the key requirements that need to be met by modern POS may include high and consistent operating speed, reliability, ease of use, remote supportability and rich functionality. Vendors and retailers are working to standardize development of computerized POS systems and simplify interconnecting POS devices. There is web based POS software that can be run on any computer with an Internet connection and supported browser, without additional software. The POS software is hosted on secure servers with real-time backups.[1, 2]

The reservation network system is when guests are referred by another hotel in the same chain or marketing group through affiliate or non-affiliate reservation network systems (Baker et al, 2000). An affiliate reservation system is a reservation system in which all hotels within the same chain participate. Guests can make reservation for accommodation at any hotel within the same group. A non-affiliate reservation network is a subscription, which is designed to connect independently operated hotels; guest can make reservation at any hotel within the same network.

When customers contact the front office with specific details of their proposed reservation, the Central Reservation Officer (CRO) checks room availability and makes reservation directly into the system. (Baker et al 2000.) Making reservation now is easier for guests with the help of computers and Internet because guests do not have to be on site before booking for a room.

Computers are widely used in front office today because of its efficiency and effectiveness in clerical, repetitive, data manipulation, number calculating, speed and accuracy. The involvement of IT does not work only in the front office but links all the departments like housekeeping, food and beverage, conference and health or leisure clubs together. The CRS connects hotels in a chain by sharing information such as available rooms, room rates and so on. Using CRS utilizes yield management to allow better and smoother control of room inventory, provide hotels with a wealth of information that could increase occupancy and revenue (Knowles, 1998). Central Reservation Officers can know the availability of rooms at a particular time of another hotel within the same chain (Baker et al, 2000).

Customers could check on marketing information, room availability and room rates with the use of the internet. The reservation clerk constantly updates information on the internet to reflect the current activities of the hotels. Through IT, automatic check-in is available to guests at the front office. Guests need credit cards to be issued with a computer-coded room key (Abbott and Lewry, 1999). The credit cards are needed to activate these machines to issue the computer-coded room key. The machine displays a menu showing the available rooms and their rate for guests to make their reservations and booking.[3]

III. METHODOLOGY

Data were collected from three chain hotels in Ghana; namely Golden Tulip, Novotel and Holiday Inn. Questionnaire (Appendix I), which included both open-end and close-end questions were developed and sent to the front office personnel in the above mentioned chain hotels. Respondents were expected to tick the appropriate check box. Unstructured interviews were also conducted with front office personnel. Some officials from the Ghana Tourist Authority (GTA) were not left out in the interview and were also made to participate in the questionnaire. A quantitative approach of data analysis was used on the data collected by means of questionnaire and interviews. Tables and charts were used to collate the data for the analysis.

Test questions such as; “Has the use of information technology made any effects on chain hotels in Ghana?” and “what significant change has information technology brought on the front office operations of chain hotels in Ghana?” were used in the research. The research was limited to only chain hotels in Ghana so as to realize the full impact.

IV. DATA ANALYSIS AND DISCUSSION OF FINDINGS

Demographic data of the respondents are analyzed here. Forty Eight questionnaires were given to Forty Eight staffs in three selected chain hotels and Ghana Tourist Authority, which was collected after a fortnight. Forty questionnaires were responded to out of the forty eight. Out of the 40 respondents, 28 were females representing 70% and 12 were males, which represent 30%. Table 1 and Figure 1 both show the number of questionnaires given to each hotel and the percentage of responded and unresponded questionnaires and Table 2 shows the age distribution of the respondents.

View on whether the use of ICT has brought any impact positively or negatively on the staffs and hotel as a whole as far as the front office operations are concerned is as shown on Table 5.

TABLE I. QUESTIONNAIRE RESPONDENT TABLE

Hotels & GTA	Responded	Not-responded	Total Questionnaire
Holiday Inn	10	2	12
Golden Tulip	12	0	12
Novotel	10	2	12
GTA	8	4	12
Total	40	8	48
Percentage	***83%***	***17%***	***100%***

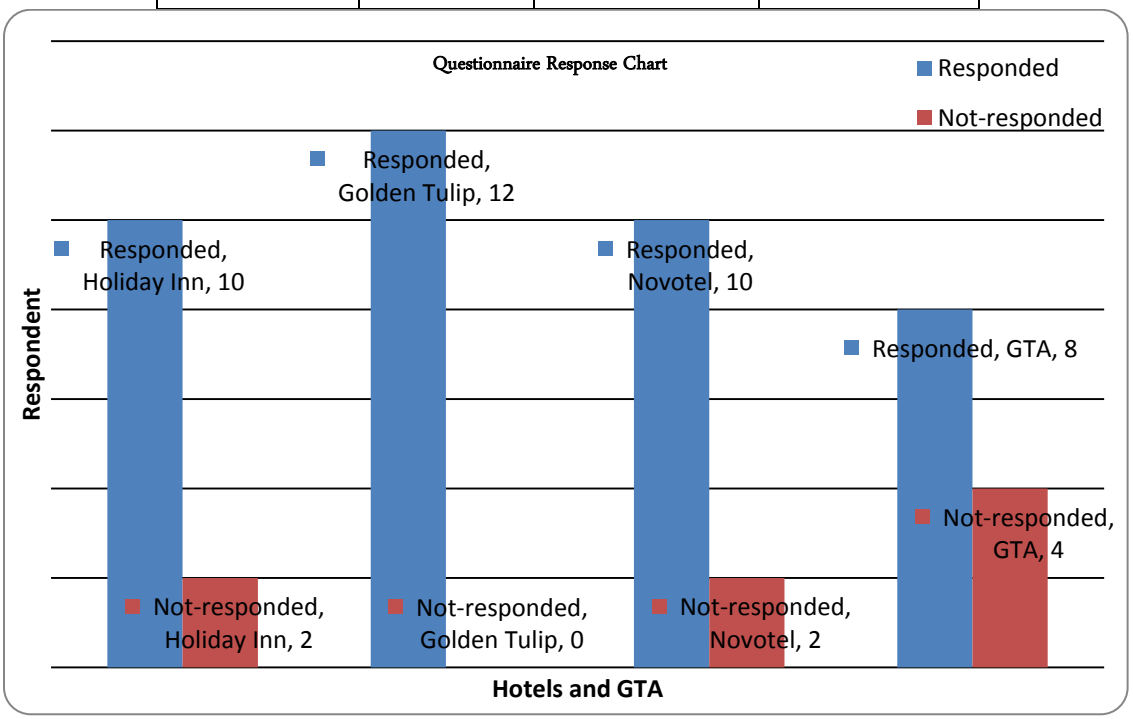


Figure 1. Questionnaire Respondent Chart

TABLE II. AGE DISTRIBUTION OF RESPONDENTS

Age (yrs)	Number of Respondents	Percentage (%)
22 - 25	12	30
26 - 30	18	45
31 - 35	6	15
36 - 40	4	10
Total	***40***	***100***

TABLE III. INTERVIEW ON IMPACT OF IT ON FRONT OFFICE OPERATIONS

Hotels & GTA	Positive	Negative	Total Interviews
Holiday Inn	4	1	5
Golden Tulip	4	2	6
Novotel	4	0	4
GTA	3	0	3
Total	15	3	18
Percentage	***83.3%***	***1.7%***	***100%***

TABLE IV. EDUCATIONAL LEVEL OF RESPONDENTS

Educational Level	Number of Respondents	Percentage (%)
Master	4	10
Degree	8	20
HND	8	20
Advance Diploma	12	30
Diploma	8	20

Total	***40***	***100***

TABLE V. IMPACT OF IT ON FRONT OFFICE OPERATIONS

Impact	Number of Respondents	Percentage (%)
Positive	35	87.5
Negative	5	12.5

V. CONCLUSION

The paper looked in totality whether the impact of using Information and Communication Technologies (ICT) in the front office operations of chain hotels is positive or negative. 87.5 per cent of the 40 respondents responded positive to the impact of using ICT in the front office of chain hotels and 12.5 per cent responded negative to the same. From Table 3, it could be seen that the respondents were carefully chosen and could say that the result in Table 4 is authentic. The use of IT has reduced queues to barely negligible at the front office during check-in and check-out because of the use of the Internet and Electronic Point-of-sale (EPOS) system that transfers instant charges on guest accounts. Guest can remotely inspect their account from the convenience of their abode.

VI. RECOMMENDATION

After careful consideration, the following recommendations could be drawn;

- The front office staff should have adequate IT training in the use of the ICT equipment.
- The front office staff must receive training to help them prevent Video Display Terminal (VDT) and Carpal Tunnel Syndrome (CTS) threats.
- Chain hotels should occasionally organize refresher courses for the front office staff to introduce them to new software and hardware applicable to their operations.
- GTA should sensitize chain hotels in Ghana on the importance of networking the operations of the hotel using ICT.
- The GTA should organize seminars and workshops at least annually on the use and importance of ICT not only the chain hotels but the hospitality industries as a whole.
- Chain hotels should encourage and promote the use of ICT in their front office operations.
- Policy makers in the tourism industry should make sure that all chain hotels in Ghana use ICT to network all their branches across the country.
- Chain hotels should encourage their guests to use the internet and telephony to make reservations.

REFERENCES

[1] Olifer, N. and Olifer, V. (2006) Computer Networks, Principles, Technologies and Protocol for Network Design, John Wiley and Sons Ltd, England.

[2] Ritchie, C. (2003) Operating Systems: Incorporating UNIX and Windows, Continuum, London and New York. 4th Edition

[3] Abbott, P. and Lewry, S. (1999) Front Office: Procedures, Social skills, Yield and Management, Elsevier, United Kingdom.

[4] Knowles, T. (1998) Hospitality Management: An Introduction, Prentice Hall, London. 2nd Edition

[5] Odom, W. (2007) CCENT/CCNA ICND1: Official Exam Certification Guide, Cisco Press, Indianapolis. 2nd Edition

[6] Lucey, T. (2005) Management Information Systems, Book Power, Britain.

[7] Baker, S., Huyton, J. and Bradley, P. (2006) Principles of Hotel Front Office Operations, Thomson Learning, London.

[8] Edexcel Limited (2010) Front Office Operations in Hospitality, BTEC Level 3 Nationals specification in Hospitality, Issue 1

[9] 1Jamel, C. (2010) Front Office Operations, Retrieved from www.satm.bilkent.edu.tr/~jamel/Foo/52148-Chapter%205.doc (23 June 2010 12:25pm)

[10] 2Kofax Front Office Server, Retrieved from http://www.kofax.com/downloads/datasheets/ds-kofax-front-office-server-en.pdf (23 June 2010 12:32pm)

[11] 3Wikipedia (2010) Hospitality Point of Sale Systems, Retrieved from http://en.wikipedia.org/wiki/Hospitality_point_of_sale_systems#Hospitality_industry (24/06/10 11:29 am)

AUTHORS PROFILE

Albert Kwansah Ansah obtained his MSc in Mobile Computing and Communications from the University of Greenwich in London, United Kingdom in 2008. He is currently a Computer Science and Engineering lecturer at the University of Mines and Technology, Tarkwa-Ghana and a member of International Association of Engineers (IAENG). His research interests include Wireless, Satellite and Mobile Communication & Application, Network Coding and Antenna Design.

Victoria S. Blankson is currently a final year BBA Hospitality Management Student in the Department of Consideration of Tourism and Hospitality, Zenith University College, Accra-Ghana. Her research interests are Front Office Operations and Hotel Management.

Millicent S. Kontoh was born in Nkawkaw, Ghana in 1982. She graduated from Tver State Technical University with Masters in Computers, Complex Systems and Networks in 2007. She is currently a lecturer at the Computer Science and Engineering Department, University of Mines and Technology, Tarkwa Ghana. M. S. Kontoh has worked in several institutions including Registrar General's Department under the Ministry of Justice and Attorney General, Ghana, Zenith and Pentecost University Colleges in Ghana. She is an associate member of the Institute of Electrical and Electronic Engineers, IEEE and served as its Student Branch Counsellor at Pentecost University, and currently an executive member of IEEE Ghana Section. Her research interests include Modelling of Software and Database Systems, Systems Analysis and Design.

APPENDIX I

The use of Information and Communication Technologies (ICT) in front office operations in chain hotels

QUESTIONNAIRE

Please tick (√) or provide the appropriate answers to the following questions where applicable.

1. Grade of the Hotel.
   .....
2. Age between: 22-25 yrs [ ] 25-30 yrs [ ] 30-40 [ ]
3. Gender: Male [ ] Female [ ]
4. Educational Level: Diploma [ ] Advanced Diploma [ ] Higher National Diploma (HND) [ ] Degree [ ] Masters [ ]
5. Marital status:
   Single [ ] Married [ ] Divorced [ ] widowed [ ]
6. How long have you been working in this hotel?
   ……
7. What is your current position?
   ……
8. How long have you been in this position?
   ……
9. How long has your hotel or office been using Information Technology?
   ……
10. Has the use of Information and Communication Technologies made any effect on your hotels and office? Yes [ ] No [ ] Please give reasons for your answer on either positive or negative effects.
    ……
    .....
11. What significant change has Information Technology brought on the front office operations or your office? Positive [ ] Negative [ ]
    Please give reasons for your answer.
    ……
    ……
12. How many times does your hotel go on training? Once a year [ ] Twice a year [ ] Thrice a year [ ]
13. What impact does the training bring on you (front office staff) and the hotel? Positive [ ] Negative [ ]
    ……
    ……
    .....
14. How is the use of Information Technology in front office operation? Easy to use [ ] Difficult to use [ ] Please give reasons for your answer.
    ……
    ……

A Digital Ecosystem-based Framework for Math Search Systems

Mohammed Q. Shatnawi
Computer Information Systems Department
Jordan University of Science and Technology
Irbid, Jordan

Qusai Q. Abuein
Computer Information Systems Department
Jordan University of Science and Technology
Irbid, Jordan

***Abstract*—Text-based search engines fall short in retrieving structured information. When searching for x(y+z) using those search engines, for example Google, it retrieves documents that contain xyz, x+y=z, (x+y+z) =xyz or any other document that contain x, y, and/or z but not x(y+z) as a standalone math expression. The reason behind this shortage; is that the text-based search engines ignore the structure of the mathematical expressions.**

Several issues are associated with designing and implementing math-based search systems. Those systems must be able to differentiate between a user query that contains a mathematical expression, and any other query that contains only a text term. A reliable indexing approach, along with a flexible and efficient representation technique are highly required. Eventually, text-based search systems must be able to process mathematical expressions that are well-structured and have properties that make them different from other forms of text.

Here, in this context we take advantage from the concept of digital ecosystems to refine the text search process so it becomes applicable in searching for a mathematical expression. In this research, a framework that contains the basic building blocks of a math-based search system is designed.

Keywords-component; digital ecosystem; math search; information retrieval; text-based search engines; structured information; indexing approach; representation technique.

I. INTRODUCTION

A mathematical expression has many equivalent expressions [1]; this makes the process of searching for a mathematical expression is different than searching for other types of information. For example, the expression x^-1 is mathematically equivalent to 1/x. Traditional search engines do not differentiate between mathematical expression and any other types of information. Google treats both expressions as text-based ones.

In fact, the mathematical expression has certain properties that make it far different from other types of information. Actually, the structure of the mathematical expressions conveys their correct interpretation [2][3][4].

II. PROBLEM STATEMENT

Currently, traditional search engines are not able to search for math expressions or even recognize math notations and symbols. Thus, to search for a certain math expression, users need to consider the following [3]:

- How to enable those search engines to recognize math symbols?
- Do those search engines understand the equivalency in math?
- Do those search engines understand the structure of math expressions?

All of the above need to be considered in order to enable those search engines to satisfy the user needs when he/she searches for math contents as well as other types of contents. The specific needs of users will be investigated in further details in the following sections.

III. RESEARCH GOAL

Building a math-based search system can be achieved using two different approaches. The first approach is to take advantage of the text-based search systems and tailor them to be adequate for math-based search quires. The second one is to build math aware search systems from scratch, based on the new emerging technologies. Either one has its pros and cons [2].

The goal of this research is to design a framework based on digital ecosystem properties [5] [6] to support math search on the web [7]. The proposed framework consists of several components that are needed to support search activities on math-based web data with a high precision. The detailed description of the proposed framework will explain all related issues of math-based search systems.

IV. ACCESSING MATH EXPRESSIONS ON THE WEB

Virtually all searches are text-based [8] [9], thus, there are problems associated with accessing math expressions on the Web. Those problems can be summarized as follows:

- Unless we have an agreed upon technique that should be understood by both users and search engines, a user needs to know the best search terms and the best way to write a query to be used in searching for any mathematical expression.
- When a user searches for a mathematical expression, there would be non-alphabetical symbols that are not understood by current search engines (e.g. Log10x+y^2).

- The same expression can be rewritten in many different, yet equivalent ways (e.g. 1/x and x^-1).
- Text-based search engines do not consider the syntax of a mathematical expression as one of its main features.
- The used approaches to search for equivalent text terms (i.e. thesaurus to search for synonyms) are not feasible for searching for an equivalent mathematical expression.

Relatively speaking, "the text is the only data type that lends itself to a full functional processing" [8].

A. Current Search Engines and Math Search Issues

Text-based search engines cannot search efficiently for different types of mathematical constructs (e.g. axioms, formulas, etc). Mathematical expressions have some distinct properties that make current search engines inadequate to search for such expressions. There are issues that the current search environment has never had to face. Three of them will be mentioned according to what authors of [2] mentioned:

- Searching for a mathematical expression is usually combined with non-alphabetical symbols (e.g. x^3 dy/dx, x**2, etc).
- Different types of mathematical constructs are structured and the structure itself conveys the meaning of these expressions.
- The more challenging issue, is that the same expression can be represented in many different ways. For example, 1/3 mathematically is the same as 3^-1.

V. Mathematical Expression As Search Terms

Mathematical expressions are a distinct type of information. Searching the Web for a mathematical expression is not a well-defined process; the result of the search is unexpected most of the time. The inaccurate result is due to the nature of the mathematical expression search process, which is not based on clear and structured rules. In addition, the available techniques are not applicable to such expressions but they are designed and tailored to work with normal text along with different kinds of documents (e.g. multimedia).

In this paper, the concentration will be on the main three components of the proposed framework, which are:

- The Mapping component
- The Representing component
- The Indexing component

VI. The Mapping Component

Theoretically, a mathematical expression may be represented in different number of ways and sometimes in infinite number of ways. Therefore, we need to come up with a reliable technique to solve out that problem by mapping the different representations into a unique format to be used during the search process thereafter.

One major problem of not being able to retrieve relevant items is the inconsistency between the author's vocabulary and the user's vocabulary. Therefore, the user may search for a term that is not provided by the author. This problem has been studied in text search, and there are some proposed solutions; such as searching for the synonyms during the search process using thesaurus lookup. A similar problem related to equivalency exists when you search for a mathematical expression, because the term y+x is the same as x+y mathematically,

Although the current search engines are equipped with tools to enhance their ability in retrieving items that contain a certain type of mathematical expressions, they still fail in retrieving the documents that contain variants of that mathematical expression. Therefore, there is a need for a way to retrieve the documents that contain, not only the expression itself, but also the expression's equivalent forms.

The online-reasoning systems can, in theory, be used to check for equivalence between query expressions and content expressions. Those systems would take prohibitively a long time to check whether a query expression is equivalent (or not) to the expressions in the contents.

Another important reason for the failure of current search engines in retrieving mathematical expressions is that search engines do not understand mathematical structures, but they well-understand text because a word in an unstructured text is simply a word with no data type definition and no conceptual definition.

Mathematical expressions are well structured, and the structure itself holds their correct interpretations. For example, in math there is a difference between 2*(x2-x3) and 2*x2-x3. However, if we were doing text retrieval there would be no difference between both expressions.

A. Definition of Mapping

Mapping is a sequence of transformations that is concerned with transforming an original expression form one algebraic/structural form into an equivalent one. According to this definition, the Mapping is divided into two types: algebraic and structural Mapping. In algebraic Mapping, the process of Mapping is done on the expression in its algebraic form. Therefore, the algebraic form changes after mapping.

The same Mapping can be called structural mapping when the structure of the parse-tree representation is changed after applying the Mapping process. In structural Mapping, the expression's parse tree [10] [11] structure will change after Mapping the mathematical expression to its equivalent one. For this reason, we call it structural.

For example, once the expression x+z+y is mapped to its equivalent form x+y+z, this mapping is called algebraic; because the algebraic form of the expression has been changed. In addition, the parse tree for the expression x+z+y, before the mapping, is shown in Figure 1.

The structure of the parse-tree representation shown in Figure 1 can be changed to the parse-tree that is shown in Figure 2

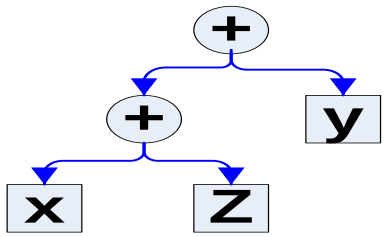


Figure 1. Parse-tree representation for x+z+y before the mapping

.

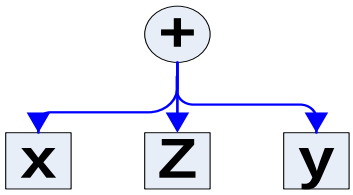


Figure 2. Parse-tree representation for x+z+y after the mapping

Notice that the structure of the parse tree has changed from one form to another after applying the mapping process.

B. Equivalence Detection and Mapping

The Equivalence Detection and Mapping (EDM) aims to transform the expression tree into a normalized one. This tree is equivalent to the original tree, but it is an agreed upon representation, based on some rules, to facilitate the search process. Therefore, the normalized tree should be the common form between the searchable database and the mathematical expression as a search term. The proposed framework should be able to update the query and the database content dynamically, so that they are both transformed into a common form.

A detailed description about this component can be found in [1], and [12]. The authors of [12] outline the main component of this subsystem. The Mapping rules are built based on a context free grammar (CFG). The authors built the component that contains the rules that are responsible for the mapping process. The mapping process verifies that a set of expressions are equivalent. For example, this component verifies that the x+y and y+x are equivalent or not. This approach is different than the theorem proving systems. The theorem proving systems verify whether the given two expressions are equivalent; whereas the EDM finds all the equivalent forms of a certain mathematical expression [13]. The process done by the EDM is faster than the one that is applied in the theorem proving systems.

In order to detect the equivalency of math expressions, a context free grammar is built to verify the format of the added mapping rules. This component is built based on the properties of the digital ecosystems [5], in which the system is able to update itself based on different user specifications, and based upon any added rules. This component is able to normalize the database content and the user query based on a list of mapping rules. There is no need to modify the system, because the mapping component is reacting automatically.

The mapping system is built based on the grammar that is responsible for verifying the format of any added mapping rule. The rules are added to check the equivalency of math expressions. The purpose of the grammar is to constraint the format of the added mapping rules. Any added rule that does not comply with the predefined grammar, is send back to the user in order to reformat it again. Notice that, the above rules can be written in a different way based on the way the user writes the grammar. Thus, the grammar decides how those rules are written and decides the further steps to be taken thereafter.

C. WildCards in Math Search Systems

The wildcards have been used in math search systems to achieve several purposes. The authors of [14] have used the wildcards to extend the current math query languages. The introduced three sets of wildcards are used for more precise structural search, and multi-level of abstraction. The authors of [14] introduced wildcards for several math operations, such as matrices, partial differentiation, and for function composition.

The query language that is introduced by the authors of [15] contains a set of wildcards. The implementation of this query language maps the queries written in that language into Xpath/Xquery quires [16][17]. The authors of [15] assumed that the math content is in MathML.

The introduced framework in this research can benefit from the proposed wildcards in [15] by providing a set of wildcards that can be used in the mapping process. In addition, the wildcards can be used during the search process in which the math query language in [14] can be tailored to be used with the proposed framework; especially the math expressions in [14] are assumed to be represented as parse-trees.

D. Generic Mapping

Based on the mapping component (i.e. the grammar), the system administrator should be able to add any valid mathematical equivalence rule. The Mapping system should be able to detect equivalency in math expressions. The rules tell whether two or more expressions are equivalent or not. In addition, the areas of math that our system has provided equivalence detection for must be determined.

Also, the system should be able to determine which group of users is targeted. Algorithms are developed to detect equivalency for any added rule that conforms to the grammar; any added rule to the generic Mapping (GM) system is derived from a general principle in which a rule is admissible, if and only if, there is a corresponding transformation on the parse-tree [12].

The GM processes a massive amount of math content. Thus, there are difficulties associated with searching such content using current search engines as mentioned before. Consequently, this research adapts the concept and properties of digital ecosystems trying to enhance the ability of GM system in increasing the precision and/or recall when searching math content.

Accordingly, the GM system has been developed to be:

- Able to be incorporated in different environments, i.e. web-based systems, math-search systems, etc.
- Designed as a separate component that can cooperate efficiently with other ecosystems.

- Flexible in which a user can choose whether to apply the GM or not. Users can notice the benefits of the GM after it has been used.
- Scalable in which the GM can be easily expanded to include all related math content.

Any added Mapping rule is validated in order to verify whether it is compliant with the grammar or not. This process is implemented using javaCC [18].

Figure 3 summarizes the detailed components of the Mapping sub-system.

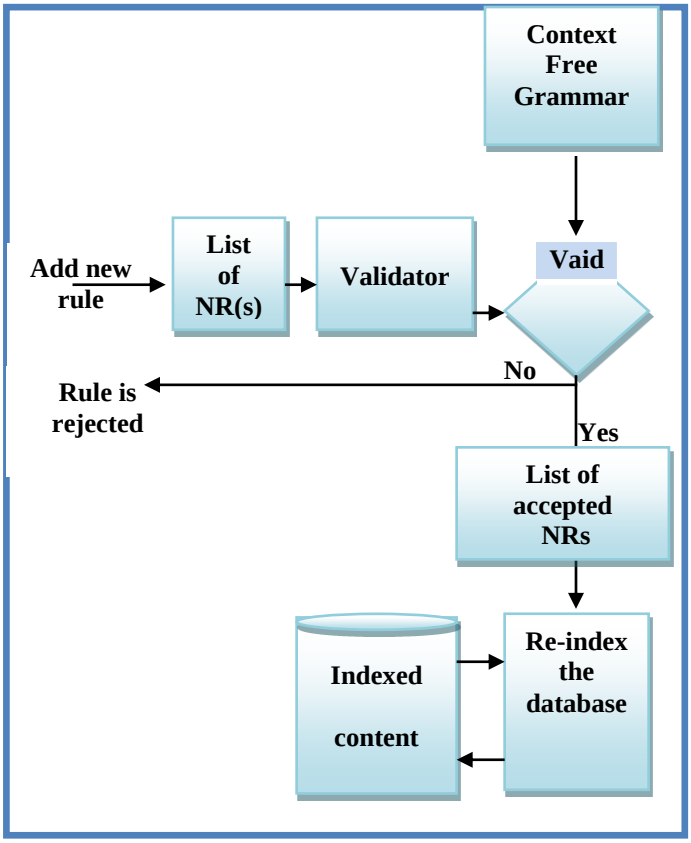


Figure 3. General Mapping System Based on a Conetxt Free Grammar

The framework that appears in Figure 3 can be enhanced by adding an intelligent component. This component can achieve several tasks on behalf of the user and the system. For example, the intelligent component can automatically check the correctness of the user's query, and at the same time, provides help for the user trying to figure out the query. The intelligent component can contain several well-known math functions and properties (axioms, trigonometric functions… etc). Notice that the context free grammar is defined by the user, in order to enforce certain format for the mapping rules. The mapping rules are responsible for mapping an expression to all of its equivalent forms.

VII. The Represening Component

Before getting into the proposed indexing approach, there is a need to discuss the different approaches to represent mathematical expressions.

Currently, there are many available representation techniques that have been used to represent mathematical expressions. For example, text-based mathematical constructs, XML-based math content, tree-based representation, and other representation techniques such as Box model [4].

In the designed framework, the parse tree representation is adapted as an efficient and reliable representation technique. For example, the tree operations are efficient; comparing sub-trees operators (i.e. sub-expressions) is easily implemented. The more important feature is that the structure of a mathematical expression conveys the correct interpretation of those expressions, and the tree representation can hold the structure of those expressions. In fact, using the parse trees to represent the mathematical expressions maintain the properties of math expressions. For example, the operator precedence is maintained when representing the expressions using parse trees.

As new math constructs are being added to the web, the use of parse trees is highly recommended. The already math-based content can be converted into their parse tree representation. The conversion process has some technical issues that can be handled once. Those different representations can be mapped to their parse trees' representation.

The sub-system that is responsible for converting the different representation into a unique one can be implemented based on the digital ecosystems properties. For example, the conversion can be done automatically based on the existing representation of an expression. The component that does the conversion can be enhanced with some rules, knowledge, or any other model that might contribute to the correct conversion.

VIII. The Indexing Component

There is not a well-defined indexing approach that can be used to index math-based content. Even with the extensive indexing of Metadata, users can only search for the math expression itself [4] [19]. A mathematical expression can be found in different equivalent forms. The existing techniques are not mature enough to fulfill the special requirements that are associated with the nature of mathematical constructs.

Based on the parse tree representation, a new promising technique has been proposed to index math-based content. The whole approach depends on assigning an agreed upon values for each parse tree node. Those values can be taken from a lookup table. Certain calculations can be performed on those values to extract a unique one to be used to index the whole parse tree. This approach is similar to the hashing technique in which the idea of the function that is used in our proposed indexing technique is similar, somehow, to the hashing function.

An ongoing experiment is being implemented to test the result of this technique. The preliminary result is impressive and the more clarification and details about this technique might be available soon.

IX. Arabic Math Expression

Thus far, most of the researches that work on math expressions are interested in English-based math expressions.

The number of researches that work on processing Arabic-based math expression is relatively few. For that reason, it is recommended in this research to work on Arabic-based math expression, and expand the framework to work for both language-based math expressions.

Arabic-based math expression maintains the same structure and math properties of the English-based math expressions. The main distinction between the both of them is in the used language.

In order to expand the framework there is a need to do further steps on the current framework. For example, the mapping rules will be different because the language is different. Accordingly, the grammar that checks on the format of the mapping rules is different as well.

Several modifications are needed in order to enable the current framework to process both Arabic and English-based math expressions.

X. The Proposed Framework

The proposed framework is depicted in Figure 4, in which the math query is transformed into its equivalent parse tree representation. The steps thereafter depend on the user demand whether to search within the un-mapped database or to search within the normalized one. Once the user choses to search within the un-mapped database; the system takes the user right to the un-normalized database. The user using this system can chose to do the Mapping, and then the query is mapped into the normalized one based on a set of predefined rules of mapping. The search process after that completes as it does in any other search systems in which the indexing process is performed on the normalized database. Notice that the mapping is performed on both, the user math query and on the searchable math content (i.e. math database). The designed framework enables the user, after applying the mapping process, to search the un-normalized database as well. In this way, the user has the ability to compare the results of the search process under different scenarios.

XI. Significant Contributions

This research makes the following significant contributions to the field of math search.

- A new indexing approach that can be utilize to index math-based content.
- A proposed approach for representing different types of math constructs and an approach to convert the already existing math constructs to the new proposed representation.
- Introducing the Arabic-based math expression processing in which the same operations of English-based math expressions can be applied on the Arabic-based ones.
- Introducing the intelligent component that can be developed and enhanced with several math-related functionalities.

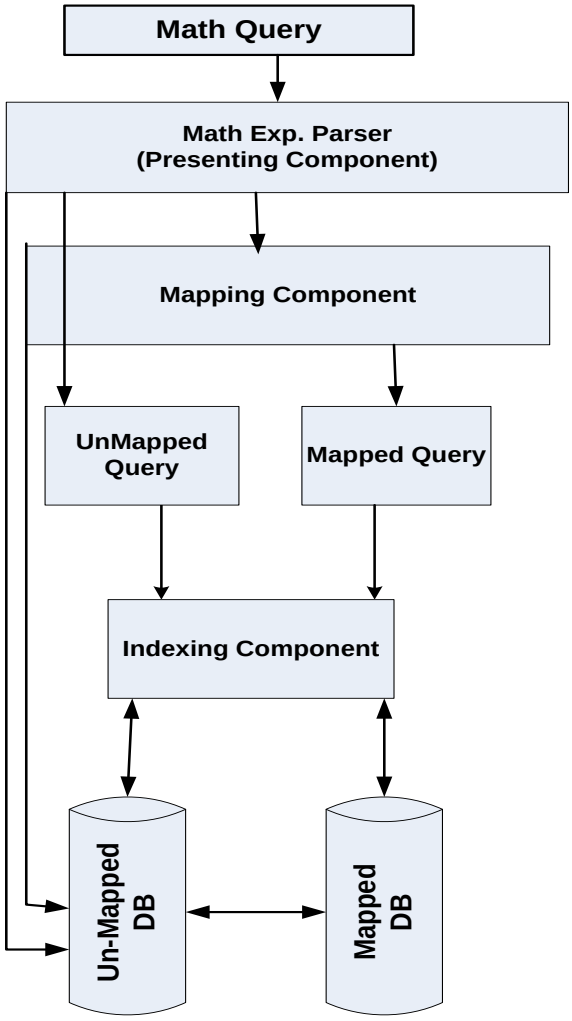


Figure 4. The proposed framework

XII. Conclusion And Future Direction

This research introduces a framework for math-based search systems based on digital ecosystems properties. The proposed framework consists of three main components.

The Mapping component shows that we have achieved some progress in searching for a mathematical expression (e.g. y+x).After applying the Mapping and equivalence rules, the recall and even precision of our search will be increased. Since we are transforming different equivalent mathematical expression into a common form, this common form will be compared against the searchable database, which contains the normalized form of that expression as well. According to that, the comparison process will end up finding most of the items that have the common mathematical expression.

The Representing component which represents the math expressions using parse tree, and convert the already existing expressions to that form. The output of this component is a math content represented using parse tree.

The reason behind this component is because the existing math content is represented in different formats, which make it more difficult to design a specialized system to search this content. Therefore, adapting the concept of digital ecosystems to standardize the way a math expression is represented minimizes the difficulties of searching such content.

The third component is the Indexing one. This component indexes the math content using an approach of assigning a unique value for each parse tree, and then uses that unique value to search for a specific tree. In addition, this approach allows for sub-expression comparison which enables the searching for a sub-expression within an expression.

The researchers still have too much to do on this field, such as:

- Combine the text-based indexing approaches with the proposed approach to a get better result when the user search for text mixed with a math expression.
- Add extra component to the Representing component (e.g. image) to be able to convert multimedia-based representations to the proposed representation.
- Expand the mapping rules, the grammar that verifies the format of the mapping rules, and the framework to work for Arabic-based math expression as well as English-based math expressions.
- Develop a comprehensive math query that can be used to search efficiently for a math expression based on a parse tree representation.

REFERENCES

[1] Mohammed Shatanwi, Abdou Youssef, "Equivalence Detection Using Parse-tree Normalization for Math Search ", ICDIM 2007 Lyon-France Oct 28-31-2007.

[2] Youssef, A. "Information Search And Retrieval of Mathematics Contents: Issues and Methods", The proceeding of the ISCA 14th International Conference on Intelligent and Adaptive Systems and Software Engineering (IASSE-2005), July 20-22, 2005, Toronto, Canada.

[3] Youssef A., "Roles of Math Search in Mathematics,"Spring-Verlag volume the Lecture Notes in Artificial Intelligence series. Also, an invited paper, the 5th Int'l Conf. on Mathematical Knowledge Management, August, 2006, UK, pp. 2-16.

[4] Abdou Youssef, Bruce R. c "Technical Aspects of the Digital Library of Mathematical Functions, Annals of Mathematics and Artificial Intelligence, Volume38, pp. 121-136, 2003.

[5] H. Boley and E. Chang, "Digital ecosystems: Principles and semantics," in Proceedings of the Inaugural IEEE International Conference on Digital Ecosystems and Technologies, 2007, pp. 398–403.

[6] P. Dini, N. Rathbone, M. Vidal, P. Hernandez, P. Ferronato, G.Briscoe, and S. Hendryx, "The digital ecosystems research vision:2010 and beyond," European Commission, Tech. Rep., 2005.

[7] Michael Kohlhase , "MATHML Presenting and Capturing Mathematics for the Web", Carnegie Mellon University, http://docbu.com/2011/09/08/mathml-presenting-and-capturing-mathematics-for-the-web/, last access in September 2010.

[8] Kowalski, Gerald J., Maybury, Mark T. "Information Storage and Retrieval Systems: Theory and Implementation", Springer, 2nd edition, 2000.

[9] Sergey Brin, Lawrence Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine", Proceedings of the 7th international conference on World Wide Web, Brisbane, Australia, 1998.

[10] Derivations and Parse Trees, http://www.cs.nuim.ie/~jpower/Courses /parsing/node24.html , last access in March 2011

[11] Christopher W. Fraser, Robert R. Henry, Todd A. Proebsting, "BURG -- Fast Optimal Instruction Selection and Tree Parsing", December 1991.

[12] Abdou Youssef, Mohammed Shatnawi, "Math Search with Equivalence Detection Using Parse-tree Normalization", The 4th International Conference on Computer Science and Information Technology, April 2006, Amman, Jordan.

[13] Automated Theorem Proving, http://www.cs.miami.edu/~tptp/OverviewOfATP.html

[14] Moody Altamimi, Abdou Youssef, "Wildcards in Math Search, Implementation Issues," 19th International Conference on Computer Applications in Industry and Engineering, November 7-9, 2007, San Francisco, California USA.

[15] Abdou Youssef (Jointly with Moody Al-Tamimi), "A more canonical form of content MathML to facilitate math search", The 2007 Extreme Markup Languages conference, Montréal, Canada, August 7-10, 2007.

[16] World Wide Web Consortium, "XML Path Language (XPath) Version 2.0," 2005. http://www.w3.org/TR/xpath20/.

[17] World Wide Web Consortium, "XQuery 1.0: An XML Query Language," 2007. http://www.w3.org/TR/xquery/

[18] Java Compiler Compiler, https://javacc.dev.java.net/, last access in February 2011.

[19] Lozier, D. W., Miller, B.R., and Saunders, B.V., "Design of a Digital Mathematical Library for Science, Technology and Education". Proceeding of the IEEE Forum on Research and Technology Advances in Digital Libraries; IEEE ADL '99, Baltimore, Maryland, May 1999.

AUTHORS PROFILE

Mohammed Q. shatnawi received his undergraduate degree in computer science from Yarmouk University/ Jordan in June 1995. Shatnawi joined the Ahli National Bank/ Amman as programmer in 1999 for 6 months. After completing his master and D.Sc. studies at the George Washington University/ DC, in January 2007 he joined the Faculty of Computer and Information Technology/ Jordan University of Science and Technology. He is currently working for the computer information systems department. His research interests are in information retrieval, supply chain management systems, CRM, data mining and algorithms.

Qusai Abuein received his B.Sc. in coputer science from Yarmouk university/ Jordan in June 1993. Abuein has joined the Yarmouk un is an assistant professor at the Computer Information Systems in Jordan University of Science and versity computer information center as a programmar and system analyst for five years. Abuein completed his master and Ph.D. in computer science from Japan and currently he is an assistant professor in Computer Information Systems Department in Jordan University of Science and Technology. His research interests are in computer cryptography, information retreival and web technolgies.

OFW-ITS-LSSVM: Weighted Classification by LS-SVM for Diabetes diagnosis

Ontological Feature weights (OFW) and intensified Tabu Search (ITS) for optimization

Fawzi Elias Bekri
Department of Computer Science & Engineering
JNTU, Hyderabad- 500 085, Andhra Pradesh, India

Dr. A. Govardhan
Professor of Computer Science & Engineering
Principal College of engineering, JNTUH, Jagityal,

***Abstract*—In accordance to the fast developing technology now a days, every field is gaining it's benefit through machines other than human involvement. Many changes are being made much advancement is possible by this developing technology. Likewise this technology is too gaining its importance in bioinformatics especially to analyse data. As we all know that diabetes is one of the present day deadly diseases prevailing. So in this paper we introduce LS-SVM classification to understand which datasets of blood may have the chance to get diabetes. Further, considering the patient's details we can predict where he has a chance to get diabetes, if so measures to cure or stop it. In this method, an optimal Tabu search model will be suggested to reduce the chances of getting it in the future.**

Keywords-machine learning; SVM; Feature reduction; feature optimization; tabu search.

I. INTRODUCTION

In the present situation we can say that diabetes has no cure. In real, it happens due to the lack of insulin which has to do along with glucose in our body. It has to be supplied into our body during loss conditions externally. Indirectly, I is the main cause for fatal heart, kidney, eye and nerve diseases, which can be overcome or prevented by good food habits and body exercises[1].

Over all this, the difficult thing is to differentiate between disease diagnosis and interpretation of diabetes data. For doing this, we are with Support vector Machine(SVM) which was developed by Vepnik[2]. Its work has been tested in many ways[3][12][4]. The utmost advantage with it is that it can even work good with nonlinear functions and it contains Radial basis Functions(RBF) which is even more precise than polynomial and linear kernel functions. The comparison of this SVM with other methods like Combined Neural Networks (CNNs), Mixture of Experts (MEs), Multilayer Perceptrons (MLPs), Probabilistic Neural Networks (PNNs) also revealed that svm methods are perfect.

In estimating the diabetes features, Feature Selection is applied. By the above analysis, if we are left with 8 features, we can come down to 4 by this feature selection. It can take out the factors not conserned with the feature set.[5][6][7][8].

PCA (Principal Component analysis) is one of the feature selection method recently gaining its importance being used in image recognision, signal processing, face recognition etc.

By applying SVM to disease datasets, it can grab a large circumference of data sets even relevant or not relevant to the diagnosing the disease. But by such features with variation, the diagnosing will not be perfect and so weighted factors are to be developed. And they were contributed by Zhichao Wang [9] giving those weights by their ontological relevance.

The LS-SVM technique at last works with 2 parameters for accurate results. Out of many datasets and values came from SVM, the choosing of 2 parameters is very important, If very high features are chosen, some datasets will be missed and if chosen with utter accuracy and care, leads to under-fitting[6,13]. So, 2 optimised solutions are to be found out possibly by Intensified Tabu Search (ITS)[14].

The working of this ITS involves 3 phases. PCA, discussed above, is to get rid of irrelevant evidences given by SVM. Then OFW is to calculate the weight of each factor which PCA thought relevant. Then comes ITS which can find out the best possible 2 parameters for SVM so that it may not be under fit or over fit.

To have a quick look on what paper contains, we shall see the initial data sets of diabetes in section II, then our 1st step of PCA reduction in section III, to weighted preferences in section IV, then OFW in section V followed by experimental results in later sections.

II. DATASET OVERVIEW

The initial data sets are gathered form UCI Machine Learning Repository[16]. It contains almost 8 categories on a whole and 768 sub categories which is really a very large database. The attributes are choose from these large data sets may be either discrete or continuous with an interval[17]. The large data base, provided now is from the following:

- Pregnant: Number of times of pregnant
- Plasma-Glucose: Plasma glucose concentration measured using a two-hour oral glucose tolerance test. Blood sugar level.
- BMI: Body mass index (w in kg/h in m)
- DPF: Diabetes pedigree function
- TricepsSFT: Triceps skin fold thickness (mm)
- Serum-Insulin: 2-hour serum insulin (mu U/mt)

- DiastolicBP: Diastolic blood pressure (mmHg)
- Age: Age of the patient (years)
- Class: Diabetes onset within five years (0 or 1)

III. FEATURE SELECTION

The 1st method which runs for reducing the data base is feature selection. The complexity of data can be reduced so that we can be left eith less datasets and can be more precise. Then comes PCA helping the classification to happen further with the help of statistical measures. The simplification of data by PCA is as follows:

D n-dimension dataset.

M principle axes a1,a2,…. These are orthogonal axes… then, covariance matrix is:

$$s = \left(\frac{1}{L}\right)\sum_{k=1}^{L}(x_k - p)^T (x_k - p)\ \ x_k \in D \quad (1)$$

Where m is the average of samples, and L is the number of samples. Therefore

$$sv_k = \lambda_k v_k \quad k \in 1,....,n \quad (2)$$

Where λ_k is the k^{th} largest Eigen value of s . The m principal components of a given sample $x_k \in D$ are given in the following

$$q = [q_1, q_2, ..., q_n] = A^T x_k \quad A = [a_1, a_2, ..., a_n]: \quad (3)$$

where $q_1, q_2, ..., q_n$ are the principal components of x_k .

LS-SVM: Of all the paper, we discussed the key idea of using SVM brought up by Vapnik[2]which plays a main role in collecting the wide database for our problem. It also has its use in solving pattern recognition and classification problems. The methods present in SVM other than polynomial and linear are its greatest assets which made it to lead global models containing structural risk minimization principle[19]. Though SVM sounds easy due to its extended results, finding the solution is difficult and what all can do is to find sparse solutions. Its difficulty arises from finding nonlinear equations. So as a solution, Suykens and Vandewalle [20] introduced least-squares SVM which results out linear equations. For the new type of SVM also the further proceeding like PCA, OFW and its usage in quantification and classification are applicable and reported in some works[23,24].

In calculation of linear equation, (y=wx+b), we use the 2 axes like regression(x) and dependent variable (y). And the best minimised cost function is

$$Q = \frac{1}{2} w^T w + \frac{1}{2}\gamma \sum_{i=1}^{N} e_i^2 \quad (3)$$

Subject to: $y_i = w^T \phi(x_i) + b + e_i \qquad i = 1,....N \quad (4)$

The formula's two parts are weight decay the 1st to generalize weights and regression error of training data is the second, whereas the parameter indicated by g is to be optimized by the user.

For a better generalization model, the most important criteria are the proper selection of features for the RBF kernel and polynomial kernel.

IV. ONTOLOGY-BASED FEATURE WEIGHTING CALCULATION

A. Feature Weight Calculation

The process of computing domain ontology feature and ontology feature weight is as follow:

a. Characteristic of the information is individually treated as a semantic category and is considered as an ontology semantic peer. The characteristics are grouped based on their semantic principles.
b. The whole relevancy of a feature is used to calculate weight of the characteristic in the ontology tree.

B. Domain Ontology-feature Graph

We construct ontology-feature graph w.r.t. a particular column of information in order to represent the domain knowledge model. There are three layers in the graph. They are:

i. Concept layer
ii. Attribute layer
iii. Data-type layer

Let us discuss them in detail,

Concept layer:

First layer has all the concepts of the ontology called ontology concept. It is explained by the attribute nodes and remaining elements of the concept layer. It can be represented as:

Ontology-Concept = {Cpt.1, Cpt.2, …., Cpt.n_{cpt}}.

For each layer, an object is considered as a node.

Attribute layer:

Second layer, Attribute layer explains the nodes in the concept layer i.e. ontology attribute following the regulations of the characteristic set.

Ontology- Attribute={Ab.1, Ab.2, ..., Ab.n_{ab}}.

Data Type Layer:

This layer explains the node of the Attribute layer following the regulations of the metadata layer i.e. Ontology-Data type.

Ontology- Data type={Dt.1, Dt.2, …, Dt.n_{dt}}

In the figure 1, the solid line shows the relative characteristics of the concept semantic layer and attribute layer whereas, dotted lines show the data type layer nodes individually.

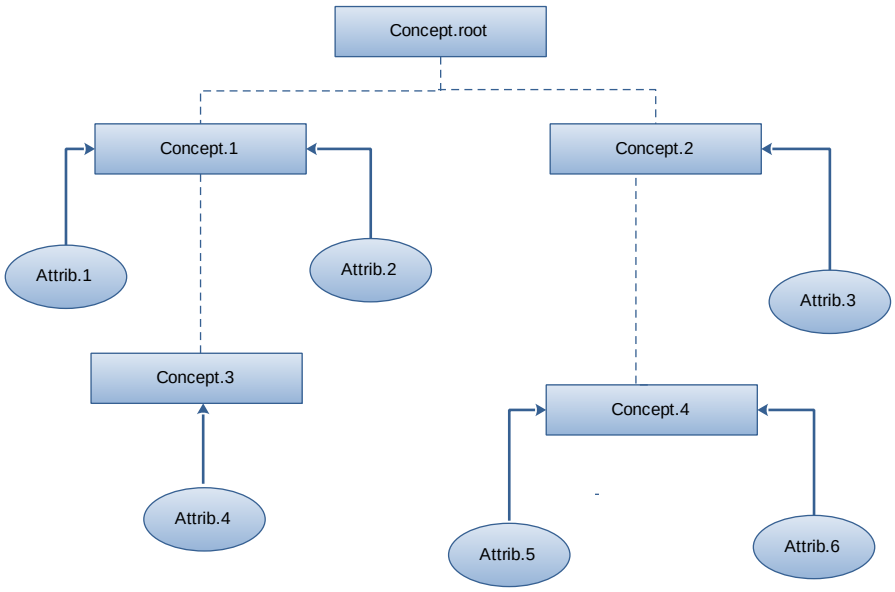


Figure 1. Ontology Feature Graph

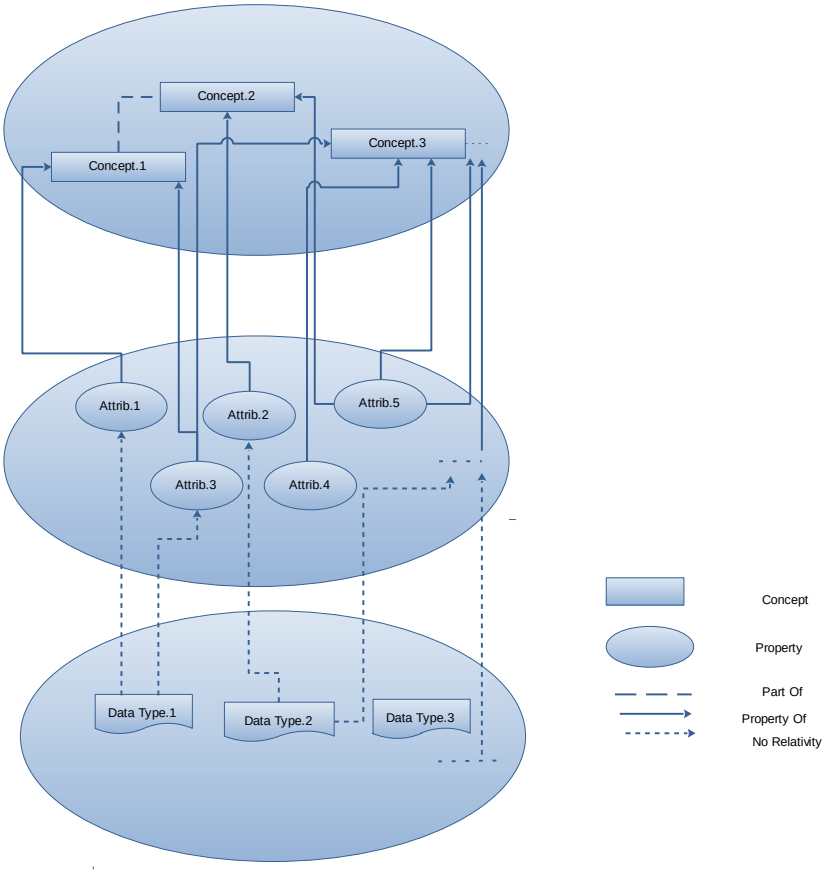


Figure 2. Ontology Feature Tree

The characteristics of the information source and database storage logical model in the domain ontology feature graph are used to compose the data object ontology node. The latter one is used to construct the nodes of the concept layer of the domain ontology characteristic graph based on its design pattern. Generally, the data object ontology node is constructed based on the remaining of the data source of the ontology graph. The database is composed by using the ontology principle as the

primary rule and the principle layer nodes of the graph which is constructed based on database.

The ontology attribute layer is constructed from the elements of the attributes which are the characteristics of the concept layer. The data type layer is used to convert the data types of the attributes to the semantic extension type. The correction between the concept layer and the attribute layer nodes can be described and computed based on the considered s of the principle layers. The communication and computation between the principle and attribute layer nodes is as following:

Semantic Ontology Correlation: The two ontology topics follow the format with predicate suchas as <Subject, Predicate, Object>. Here, Predicate represents the group of predicates,

Predicate= { partOf ,propertyOf },

It is mainly utilized in explaining the ontology predicate. Ontology relations, Concept-Concept and Concept-Attribute, are refered to as the CC and CA following the predicate set explanation.

CC stands for CC =<Concept.i, partOf, Concept.j>

CA stands for CA = <Concept.i, propertyOf, Attribute.k>

Concept.i,Concept.j ∈ Ontology-Concept;

Attribute.k ∈ Ontology-Attribute.

$$i, j < n_{cpt}, k < n_{ab}$$

Basic assumption: (Ab.i1, Ab.i2, … , $Ab.in_i$), Concept.j (Ab.j1,Ab.j2, … , $Ab.jn_j$).

$$n_i, n_j < n_{ab}.$$

The primary relation between Concept.i and Concept.j is Correlation (Concept.i,Concept.j):

$$correlation(concepti, conceptj) = \frac{|\ concepti(Ab.in_i) \bigcap conceptj(Ab.j1,\ j2,....Ab, in_j)\ |}{|concepti(Ab.in_i) \bigcup\ conceptj(Ab.j1,\ j2....Ab, in_j)|}$$

Domain Ontology Feature Tree:

It is mainly used to refer to the relationship among the nodes of the attribute layer, concept layer and their characteristics (represented in the domain ontology characteristic graph). We also make use of it in the computation of correlation between ontology-concepts and ontology-attributes.

Domain ontology characteristic tree can be referred to with the triples as

Ontology-Tree, Ontology-Tree = {<Cpt.root,partOf,Cpt.1>, <Cpt.root,partOf,Cpt.2>, <Cpt.1,propertyOf, Ab.1>, ……}

Here, in the tree, the final node i.e. the leaf node is one of the characteristics of the domain ontology whereas the branches can be represented by the concept. Thus, the discussed correlation can be calculated as:

$$Correlation(Ab.i, Ab.j) = \frac{Height(Ab.i) + Height(Ab.j)^{*}\alpha + Datatype(Ab.i, Ab.j)^{*}\beta}{(Dis\tan ce(Ab.i, Ab.j) + \alpha) + 2^{*} MAX(Height(Ab.i), HeightAb.j)) + \beta}$$

Where,

Height(Ab.i),Height(Ab.j) refer to characteristics hierarchy Ab.i and Ab.j of the concerned tree.

Boolean function Data Type(Ab.i, Ab.j)is used to compare data types of features Ab.i and Ab.j.

Distance(Ab.i, Ab.j) is the shortest path to the elements.

Max(Height(Ab.i),Height(Ab.j)) refers to the maximum length of the tree.

α,β represent the variable parameters with 0< α,β<1.

α maintains the Height and Distance ratio;

β is used in type conversion.

Here, in figure 2, the relation between Attrib.4 and Attrib.3 is computed. Height(Ab.4) = 3.

Height(Ab.3) = 2.

As we can see that both Ab.4 and Ab.3 differ in data type, Data Type(Ab.4,Ab.3)=0. Distance(Ab.4, Ab.3) = 5.

$$Correlation(Ab.4, Ab.3) = (5^{*}\alpha)/(5+\alpha+2^{*}3+\beta) = (5^{*}\alpha)/(11+\alpha+\beta)$$

Similarly, the relation between Ab.4 and Ab.2 is also computed.

Height(Ab.4) = 3;

Height(Ab.2) = 2;

As we can see that both Ab.4 and Ab.3 differ in data type, Data Type(Ab.4,Ab.2)=0. Distance(Ab.4, Ab.2)=3.

*So, Correlation(Ab.4, Ab.2) = (5*α)/(3 +α+2*3+β) = (5*α)/(9+α+β).*

Thus, we can say that Correlation(Ab.4, Ab.2)> Correlation(Ab.4,Ab.3). Hence, Ab.4 and Ab.2 are more similar. The results might change with the variables but the actual one doesn't change.

This shows that, the values of the arbitrary parameters remain unaffected over the relation among attributes. However, we can better the situation by choosing proper parameters through various tests. Thus, the formula to compute the weight of a characteristic can be drawn from the above co relations as,

$$Weight(Ab.k) = Average\sum_{i=1}^{m} correlation(Ab.k, Ab.i)$$

OFW-LSSVM

According to the Conventional LS-SVMs, the given function is performed by the equal contributions from all the

characteristics. But, actually, the various characteristics play different roles with various weights. Thus, different contributions from different characteristics can be performed by using the theory proposed by Zhichao Wang [9].

Given,

$\{x_i, y_i\}_{i=1}^{N}$

Represents coaching group and

$\alpha \in R^d$, where α represents the weighted vector.

$$\sum_{i=1}^{d} \alpha_i = 1, \quad \alpha_i \geq 0 \quad \text{-----(8)}$$

Now, the equation (3) can be used to provide optimal solution to the problem (4), which is as follows:

$$\min \frac{1}{2} \| w \|^2$$

$$s.t. y_i (w.diag(\alpha))$$

Where,

$$diag(\alpha) = \begin{pmatrix} \alpha_1 & 0\ldots & 0 \\ 0 & \alpha_2 \ldots & 0 \\ \ldots & \ldots & \ldots \\ 0 & 0\ldots & \alpha_d \end{pmatrix}$$

Substituting (8) and (9) into (5), yields the following new optimization problem:

$$\min_{w,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} \xi_i$$

$$sty_i (w.diag(\alpha) x_i + b) \geq 1 - \xi_i, \geq 0, i = 1, 2, ..., n$$

$$\sum_{i=1}^{d} \alpha_i = 1, \alpha_i \geq 0$$

Thus, we can write the categorization decision method is:

$$f(x) = sign(\sum_{i=1}^{N} L_i y_i K'(X_i, X_j) + b)$$

$K'(X_i, X_j)$ represents the weighted characteristic of the RBF kernel as

$$K'(X_i, X_j) = \exp(-\gamma \sqrt{\sum_{k=1}^{d} \alpha_k (x_{ik} - x_{jk})^2})$$

Intensified Tabu search (ITS) for characteristic selection

As we know, BDF chooses the characteristics based on the betterment of the recognition rate. After considering we came to know that BDF increases the count of support vectors according to the size of the problem [26]. This feature seems to be interesting in producing quick and better decision method, but it applies only if it is connected to the betterment in the recognition rate.

The support vectors and some more characteristics care mainly used to provide a quick and better SVM BDF. Due to this reason, in order to solve the conflict between the complexity and performance, Decision Function Quality (DFQ) criterion is used in association with regularization theory. Thus, SVM makes sure to coach right from the basic i.e. tiny dataset St', where it stands for the primary coaching group St. It will reduce the ambiguity related to the BDF. Even the primary set is also further optimized by using LBG algorithm based on certain assumptions. The basic assumption is to consider parameter k as a variable of problem in choosing model. It is so, because k may not be able to handle all kinds of prototypes generated by LBG algorithm during the process.

Hence, the value of k (i.e. the range of optimization), the characteristic subgroup β, the regularization constant C along with attributes of the kernel such as (σ with gaussian kernel) must be selected for every kernel method K using the model selection method. If we consider θ as a model, k_θ, β_θ, C_θ and σ_θ respectively will be the representatives of the attributes discussed so far. Moreover, q(θ) represents the DFQ criterion for a model θ (c.f. Section 3.1).

The Section 3.3 deals with the presumption of DFQ criterion along with a learning set S_l showing q(θ) ≡ SVM-DFQ(θ, S_l) which is to be optimized for model θ. The optimizing θ^* for q(θ) not being tractable, we decide to define a TS function for choosing a model with optimal intensification and diversification methodologies.

Decision Function Quality(DFQ):

For smooth calculation of the equation, we need DFQ for the theta we have. It can be known by the recognition rate RR with the help of complexity CP of decision function hu. Here comes the q(θ) = $R_R(h_\theta) - C_p(h_\theta)$ be the DFQ[25].

Here, the correct and accurate result from equation can be calculated by using smoothness term and fitting term in terms of recognition rate (RR).Cp indicates the smoothness term. The model complexity of a SVM BDF[25] is given by

$$C_p(h_\theta) = C_{p1} \log_2(\eta sv) + C_{p2} \log_2(\cos t(\beta)) \quad (5)$$

To discuss the parameters of the function, cp1 and cp2 are tradeoff between classification rate improvement and complexity reduction. Beta is a Boolean vector with n size of represented features. Ki is to represent cost for ith feature cost (beta) combined to the subset of selected features is:

cost (β) = $\sum \beta_i k_i$. When those costs are unknown, k_i = 1 is used for all features.

Simplification Step:

Reducing training set size is the simplest way to reduce complexity of SVM. This LBG algorithm [25] is used to simplify the dataset . The simplification details are in the below table and can be used in the further discussion:

Simplification(S,k)
S′ ⇐ ∅
FOR c ∈ {−1, +1}
T = {x | (x, c) ∈ S}
IF 2^k < |T| THEN T′ ⇐ LBG(T, k)
ELSE T′ ⇐ T
S′ ⇐ S′ ∪ {(x, c) | x ∈ T′} ENDFOR
RETURN S′

TABLE I. SYNOPSIS OF SIMPLIFICATION STEP

DFQ estimation

The Decision Function Quality (DFQ)[25] criterion of a particular model θ is calculated from a attained dataset Sl. we can observer the elocation of values from the details given in the Table 3 .Let s_t , s_v represents the datasets produced in a random split (Split function in synopsis SVM-DFQ) with | s_t | = $\frac{2}{3}S_l$, | s_v | = $\frac{1}{3}S_l$. s_t , s_v will be signifying the databases utilized to train SVM (training dataset) and to identify rate consideration (validation dataset). This dissociation is important in order to overcome the risk of over fitting when empirical estimation is used. The SMO algorithm version of the Torch library [31] is used to realize SVM training step. When SVM training is per-formed with unbalanced class datasets, it is more suitable to use Balanced Error Rate (BER) instead of classical Error Rate for the estimation of recognition rate. Recognition rate formulation (noted R_R) in Table 2 corresponds to BER estimation where m_y represents the number of examples in each class(y ∈ {+1,−1}) and $m_y^{correct}$ the number of examples correctly identified. . The kernel functions k_β utilized for training SVM are decided from a distance $d_\beta : d_\beta(x_i, x_j) = \sqrt{\sum_{l=1}^{n} \beta_l (X_i^l - x_j^l)^2}$. Utilizing d_β in the kernel function, the feature selection problem is embedded in the model selection problem. In the present study Gaussian kernels $K_\beta^G = \exp(\frac{-d_\beta^2}{\lambda_1^2})$ are utilized.

SVM-DFQ(θ,Sl)
(s_t , s_v) ⇐ Split(S_l)
$S_t^{'}$ ⇐ Simplification(s_t ,, k_θ)
h_θ ⇐ Training SVM($S_t^{'}, k_{\beta\theta}, c_\theta, \sigma_\theta$)
($m_{-1}^{correct}$, $m_{+1}^{correct}$) ⇐ Testing BDF(h_θ , S_v)
$R_R \Leftarrow \frac{m_{-1}^{correct}}{2_{m-1}} + \frac{m_{+1}^{correct}}{2_{m+1}}$ c_p ⇐ Complexity(h_θ)
q(θ) ⇐ $R_R - c_p$

TABLE II. SYNOPSIS OF DFQ CALCULATING FOR A DEFINED MODEL θ

V. FEATURE OTIMIZATION

Tabu Search specification

The main function q to be obtained produces the quality of the BDF h_θ . The main issue is to select an optimal model (good sub-optimal solution to be exact) θ^* for a function q when C_{p1} and C_{p2} are affixed. A model θ can be denoted by a set of n′ integer variables θ = (θ_1 , . . . , $\theta_{n'}$) = (β_1 , . . . , β_n , k,C′, σ′). Notations k_θ , β_θ , C_θ , σ_θ correspond respectively to k, (β_1 , . . . , β_n), $\sqrt{2}^{C'}$ and $\sqrt{2}^{\sigma'}$ in that integer representation of θ model. One basic move in our TS method corresponds to adding δ ∈ [−1, 1] to the value of a θ_i , while preserving the constraints of the model which depend on it (i.e.∀i ∈ [1, . . . , n′], θ_i ∈ [min(θ_i), . . . ,max(θ_i)] where min(θ_i) and max(θ_i) respectively denote lower and upper

bound values of θ_i variable). Above all the list of all possible neighborhood solutions is added. Among these possible solutions, the apt DFQ that is not tabu is selected. The set of all θ_{tabu}^{it} solutions θ which are tabu at the it repeated step of TS is defined as follows: θ_{tabu}^{it} = {θ ∈ Ω | ∃ i, t′ : t′ ∈[1, . . . , t], $\theta_i \neq \theta_i^{it-1} \wedge \theta_i = \theta_i^{it-t'}$ } with - the set of all solutions and t an adjustable parameter for the short memory used by TS (for experimental results t = $\sum_{i=1}^{n'} \max(\theta_i) - \min(\theta_i)$. The idea is that a variable θ_i could be changed only if its new value is not present in the short memory. Then, our TS method does not go back to a value of θ_i previously changed in short time, avoiding by that mechanism undesirable oscillation effects. Tabu status of solutions θ_{tabu}^{it} may prohibit some attractive moves at iteration it. Therefore, our TS uses an aspiration criterion which consists in allowing a move (even if it is tabu) if it results in a solution with an objective value better than that of the current best-known solution.

The initialization of model θ with our TS model selection is the following:

- $K_\theta - [\log_2(\max(m_{+1}, m_{-1}))/3)$
- C_θ = 1 and σ_θ = 1,
- ∀ i : β_i = 1.

In the present formula K_θ , m_{+1} and m_{-1} denotes positive and negative classes in binary sub-problems. The value of K_θ permits to begin with enough minimum datasets to get low training times with SVM for the first step.

Using intensification and diversification strategies develops TS methods [30]. The selected model should handle two kinds of problems. The first problem is testing all moves between two repetitions with a great number of features which is time-consuming. Especially, it is a waste of time to investigate moves which are linked to features where real solution is not suitable. Thus, emphasizing on moves which are only linked to SVM hyper parameters or simplification level is better than to discover new solutions. Coming to second problem, it is difficult for TS method to free from deep valleys or big clusters of poor solutions by using the short memory which effect in not tab solutions. Utilizing diversified solutions helps in win over of the problem. This is handled by enlarging step size (δ > 1) of moves and by pointing the use of all types of moves (except feature selection moves for the reason stated above).In present TS method, intensification and diversification strategies are utilized one by one and start with the intensification strategy. Later on we deal about the two strategies.

Intensification strategy

In the intensification algorithm synopsis of Table 4, Extensive Search survey all possible basic moves, whereas Fast Extensive Search explores only eligible basic moves which are not related to feature selection (i.e. changing the value of β). $\eta_{promising}$ Controls when the real solution is seen as enough and this one allows switching between the two functions mentioned.

BestNotTabu correlate to the move procedure chosen in the above part (the best tabu solution is chosen if all moves are tabu). In this synopsis, $\theta_{intensification}$ corresponds to the best solution found into a same phase of intensification, although $\theta_{best-known}$ corresponds to the best solution found in all intensification and diversification steps.

Nmax is the maximum number of intensification redundancy for which no development of the last best intensification solution $(\theta_{intensification})$ are identified as failure of the intensification strategy. Nfailure counts the number of failures of intensification strategy.

If Nfailure is higher than a fixed maximum number of failures max then ITS method stops and returns the solution $\theta_{best-known}$. If a solution in £ next has a QDF which is better than $\theta_{best-known}$, aspiration mechanism is used. That solution is selected as the new $\theta_{best-known}$ and $n_{failure}$ is reset to zero.

Diversification strategy

In the diversification algorithm synopsis of Table 5, suitable variable (one which does not have a link with features) is selected (Select Eligible Variable) by random and a jump of ±δ is performed by modifying the chosen variable in the real solution.

There are only two investigated moves (Two Move) to force the diversification of identified solutions. The jump size enlarges with the number of successive failures ($n_{failure}$) of the intensification strategy to investigate more different regions.

In the process of the diversification redundancy, the best visited solution is saved $\theta_{devertification}$ and chosen as the bening solution for the next intensification step $(\theta^{it}{}_{intensification} = \theta^{it-1}{}_{deversification})$. In the TS investigation, when aspiration is included, the strategy automatically moves to intensification and the number of failures is rearranged ($n_{failure} = 0$).

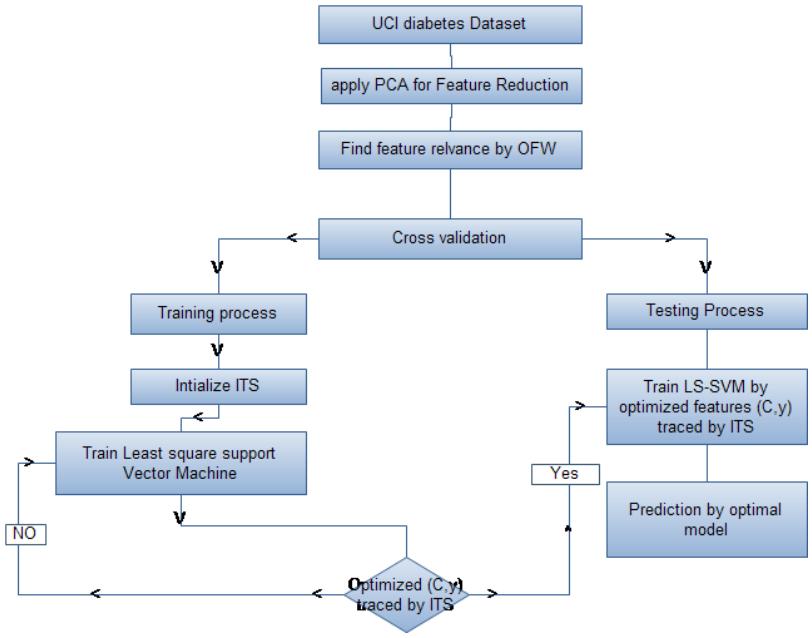


Figure 3. Flowchart of the OFW-ITS-LSSVM

VI. Proposed Method Ofw-Its-Lssvm

This part explains the desired method (OFW-ITS-LSSVM) for the identifying of diabetes diseases (see figure3). Especially the system works in three stages automatically

1) PCA is applied for feature reduction
2) Best feature weights are estimated using OFW
3) ITS is employed for finding the optimal values for C and γ.

At first, PCA method is used to identify four features from diabetes dataset. Thus, in feature choosing stage, only large principal components will be utilized. Then, the OFW-LSSVM is used to classify patients, the feature weights which are received by OFW and at last , the MCS algorithm is used to detect the best value for C and γ parameters of OFW-LSSVM. The description of training procedure is:

1. Set up parameters of ITS and initialize the population of n nests (Algorithm 1)
2. Compute the corresponding fitness function formulated by $\frac{classified}{total}$ (total denotes the number of training samples, and classified denotes the number of correct classified samples) for each particle.
3. 3. Find the best solution using ITS

VII. Experimental Results

The OFW-ITS-LSSVM model was compared with other popular models like LS-SVM, PCA-LS-SVM, PCA-MI-LS-SVM, MI-CS-SVM and PCA-PSO-LS-SVM classifiers. We utilized fold cross validation develop the holdout method. The data set was divided into k subsets, and the holdout method was iterated k times.

Every time, one of the k subsets is utilized as the test set and rest are put together to form a training set. Then the average error across all k trials is computed (Polat and Günes, 2007). This method was used as 10 -fold cross validation in our experiments. We considered the related parameters of PSO in PCA-PSO-LS- SVM classifier as follows: swarm size was set to 50; the parameters C and γ were arbitrary taken from the intervals $[10^{-3}, 200] and [10^{-3}, 2]$, respectively.

The inertia weight was 0.9, acceleration constants C_1 and C_2 was fixed to 2, and maximum number of redundancy was fixed to 70. Classification results of classifiers were shown in a confusion matrix. Like displayed in table 4, each cell has the raw number of examples categories to correspond intergration of real system results.

Output/desired	Non-diabetic	Diabetic	Method
Non-diabetics	44	6	LS-SVM (Polat et al., 2008)
Diabetics	11	17	
Non-diabetics	45	5	PCA-LS-SVM
Diabetics	9	19	
Non-diabetics	44	6	PCA-MI-LS-SVM
Diabetics	4	24	
Non-diabetics	45	5	PCA-PSO-LS-SVM
Diabetics	4	24	
Non-diabetics	48	2	MI-MCS-SVM
Diabetics	3	25	
Non-diabetics	49	1	OFW-ITS-LS-SVM
Diabetics	1	27	

TABLE III. Confusion Matrix

Thus it shows the frequency of disease how a patient is misclassified. Furthermore, Table 5 displays the categories accuracies of OFW-ITS-LSSVM. The present model gets the correct categories accuracy of 95.78% among classifiers on the test set. Determining the test performance of the classifiers is done by addition of specificity and sensitivity that are classified as: Specificity: number of true negative decisions / number of real negative case sensitivity: number of true positive decisions / number of real positive cases.

A true positive decision happens only if the positive expectation of the network mingles with a positive expectation of the physician. A true negative decision happens if the two i.e. network and the physician advice negative expectation.

Methods	Sensitivity (%)	Specificity (%)	Classification accuracy
LSSVM [32]	73.91	80	78.21
LSSVM with PCA [33]	79.16	83.33	82.05
LSSVM with MI and PCA[2]	80	91.66	87.17
LSSVM with PCA and PSO[2]	82.75	91.83	88.46
SVM with PCA, IM and MCS[2]	92.59	94.11	93.58
OFW-ITS-LSSVM	94.96	97.76	95.78

TABLE IV. THE VALUES OF THE STATISTICAL PARAMETERS OF THE CLASSIFIERS

As per the Table 6, it is observed that utilizing the LSSVM classifier with OFW and ITS, it is easy to get the correct classification accuracy compared to other methods. Hence it is apt to say that this method gives a high rate of accuracy in identifying of Diabetes disease. The method can also combine with software to help the physicians to take final decision confidently.

Method	Classification accuracy
QDA	59.5
C4.5 rules	67
RBF	68.23
C4.5 (5xCV)	72
Bayes	72.2
Kohonen	72.8
ASR	74.3
DB-CART	74.4
Naïve Bayes	74.5
CART DT	74.7
BP	75.2
SNB	75.4
NB	75.5
kNN	75.5
MML	75.5
RBF	75.7
LVQ	75.8
Semi-Naïve Bayes (5xCV)	76
MLP + BP	76.4
FDA	76.5
ASI	76.6
SMART	76.8
GTO DT (5xCV)	76.8
BFGS quasi Newton	77.08
LM	77.08
LDA	77.5
GD	77.6
SVM (5xCV)	77.6
GDA-LS-SVM	79.16
GRNN	80.21
LDA-MWSVM	89.74
MI-MCS-SVM	93.58
OFW-ITS-LSSVM	95.78

TABLE V. CLASSIFICATION ACCURACY: COMPARING OFW-ITS-LSSVM WITH OTHER METHODS FROM LITERATURE

VIII. CONCLUSIONS

Over all the work propose a new automatic method to diagnose Diabetes disease depend on Feature Weighted Support Vector Machines and Modified Cuckoo Search. For discarding the other features, Principal Component Analysis was utilized. Later Mutual Information was used to the chose features to weight them depend on their related task of classification. Outcome proves that it devises the accuracy of the method. In addition to, Modified Cuckoo Search is utilized that allows the quick change of the algorithm and locate the correct values for parameters of SVM. The outcome has proved that the present model is faster and significantly more reliable than other models.. The method can also combined with software to help the physicians to take final decision confidently in order to diagnose Diabetic disease.

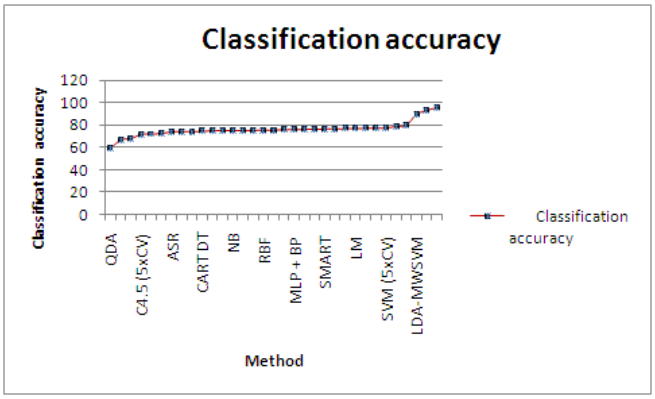


Figure 4. Line chart representation of comparing OFW-ITS-LSSVM with other methods from literature

REFERENCES

[1] [1] Polat, K., Güneş, S., 2007. An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease. Digital Signal Processing 17, 702-710.

[2] Vapnik, V., 1995. The Nature of Statistical Learning Theory, New York.

[3] Çalişir, D., Doğantekin, E., 2011. An automatic diabetes diagnosis system based on LDA-Wavelet Support Vector

[4] Übeyli, E.D., 2007. Comparison of different classification algorithms in clinical decision-making. Expert Systems 24, 17-31.

[5] Acır, N., Özdamar, Ö., Güzeliş, C., 2006. Automatic classification of auditory brainstem responses using SVM-based feature selection algorithm for threshold detection. Engineering Applications of Artificial Intelligence 19, 209-218.

[6] Lin, M., Oki, T., Holloway, T., Streets, D.G., Bengtsson, M., Kanae, S., 2008. Long-range transport of acidifying substances in East Asia—Part I:

model evaluation and sensitivity studies. Atmospheric Environment, in press, doi:10.1016/j.atmosenv.2008.04.008.

[7] Valentini, G., Muselli, M., Ruffino, F., 2004. Cancer recognition with bagged ensembles of support vector machines. Neurocomputing 56, 461-466.

[8] Zhang, Y.L., Guo, N., Du, H., Li, W.H., 2005. Automated defect recognition of C- SAM images in IC packaging using Support Vector Machines. The International Journal of Advanced Manufacturing Technology 25, 1191-1196.

[9] Lei Zhang , Zhichao Wang "Ontology-based clustering algorithm with feature weights",2010Journal of Computational Information Systems 6:9 (2010) 2959-2966.

[10] Karabatak, M., Ince, M.C., 2009. An expert system for detection of breast cancer based on association rules and neural network. Expert Systems with Applications 36, 3465-3469.

[11] Mehmet Fatih, A., 2009. Support vector machines combined with feature selection for breast cancer diagnosis. Expert Systems with Applications 36, 3240-3247.

[12] Polat, K., Güneş, S., Arslan, A., 2008. A cascade learning system for classification of diabetes disease: Generalized Discriminant Analysis and Least Square Support Vector Machine. Expert Systems with Applications 34, 482-487.

[13] Pardo, M., Sberveglieri, G., 2005. Classification of electronic nose data with support vector machines. Sensors and Actuators B: Chemical 107, 730-737.

[14] Fred Glover, Tabu search fundamentals and uses, http://leeds-faculty.colorado.edu/glover/TS%20-%20Fundamentals&Uses.pdf, 1995

[15] Xing, H.-j., Ha, M.-h., Hu, B.-g., Tian, D.-z., 2009. Linear feature-weighted support vector machine. Fuzzy Information and Engineering 1, 289-305.

[16] Asuncion, A., Newman, D. J. (2007) Pima Indians Diabetes Data Set, UCI Machine Learning Repository, http://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabets, Irvine, CA: University of California, School of Information and Computer Science.

[17] Cios, K. J., Pedrycz, W., Swiniarski, R.W., Kurgan, L. A. (2007) Data Mining: A Knowledge Discovery Approach, New York: Springer.

[18] Vapnik, V.; Statistical Learning Theory, John Wiley: New York, 1998.

[19] Sun J, Xu W, Feng B, A Global Search Strategy of Quantum- Behaved Particle Swarm Optimization. In Proc. of the 2004 IEEE Conf. on Cybernetics and Intelligent Systems, Singapore: 291 – 294, 2004.

[20] Suykens, J. A. K.; Vandewalle, J.; Neural Process. Lett. 1999, 9, 293.

[21] Suykens, J. A. K.; van Gestel, T.; de Brabanter, J.; de Moor, B.; Vandewalle, J.; Least-Squares Support Vector Machines, World Scientifics: Singapore, 2002.

[22] Zou, T.; Dou, Y.; Mi, H.; Zou, J.; Ren, Y.; Anal. Biochem. 2006, 355, 1.

[23] Ke, Y.; Yiyu, C.; Chinese J. Anal. Chem. 2006, 34, 561.

[24] Niazi, A.; Ghasemi, J.; Yazdanipour, A.; Spectrochim. Acta Part A 2007, 68, 523.

[25] Varewyck, M.; Martens, J.-P.; , "A Practical Approach to Model Selection for Support Vector Machines With a Gaussian Kernel," Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on , vol.41, no.2, pp.330-340, April 2011 doi: 10.1109/TSMCB.2010.2053026

[26] I. Steinwart. Sparseness of support vector machines - some asymptotically sharp bounds. In NIPS, pages 169–184, 2004.

[27] A. Tikhonov and V. Arsenin. Solution of Ill-posed Problems. Winston & Sons, 1977.

[28] A. Tikhonov and V. Arsenin. Solution of Ill-posed Problems. Winston & Sons, 1977.

[29] A. Tikhonov and V. Arsenin. Ill-Posed Pproblems: Theory and Applications. Kluwer Academic Publishers, 1994.

[30] F. Glover and M. Laguna. Tabu search. Kluwer Academic Publishers, 1997.

[31] R. Collobert and S. Bengio. SVMTorch: Support vector machines for large-scale regression problems. In Journal of Machine Learning Research, volume 1, pages 143–160, 2001.

[32] Least Squares Support Vector Machines for Classification and nonlinear modelling PASE 2000 (2000) by J. A. K. Suykens posted to classification lssvm pattern_recognition regression svm by Borelli on 2006-01-18

[33] Davar Giveki, Hamid Salimi, GholamReza Bahmanyar, Younes Khademian, Automatic Detection of Diabetes Diagnosis using Feature Weighted Support Vector Machines based on Mutual Information and Modified Cuckoo Search, arXiv:1201.2173v1, ARXIV, 01/2012

AUTHORS PROFILE

Fawzi Elias Bekri

He studied B.Sc IT and M.Sc IT at sikkim manipal University, Manglore. He did his M.Phil at JNTU, Hyderabad. Now he is doing Ph.D at Jawaharlal Nehru Technological University (JNTU), Hyderabad, A. P., India. His areas of interest include Data mining, KDD in healthcare sector, Software Engineering, Databases and Object Oriented Technologies.

Dr.A.Govardhan

Received Ph.D. degree in Computer Science and Engineering from Jawaharlal Nehru Technological University in 2003, M.Tech. from Jawaharlal Nehru University in 1994 and B.E. from Osmania University in 1992. He is working as a Principal of Jawaharlal Nehru Technological University, Jagitial. He has published around 108 papers in various national and international Journals/conferences. His research of interest includes Databases, Data Warehousing & Mining, Information Retrieval, Computer Networks, Image Processing, Software Engineering, Search Engines and Object Oriented Technologies.

OCC: Ordered congestion control with cross layer support in Manet routing

T.Suryaprakash Reddy
Professor
Department of CSE
Krishna Chaitanya Institute of Technology & Sciences
Markapuram,Prakasam(dist),A.P,INDIA.

Dr.P.Chenna Reddy
Associate Professor and HOD
Department of CSE
JNTU College of Engineering
Pulivendula,Kadapa(dist),A.P,INDIA

***Abstract*—In the recent times many accessible congestion control procedures have no capability to differentiate involving two major problems like packet loss by link crash and packet loss by congestion. Consequently these resolutions effect in form of wastage of possessions because they target only on the packet drop by link crash that has a needless importance. Consumption of energy and possessions in order to make the basis node attentive regarding the congestion occurring in routing path is the supplementary drawback in most of the accessible procedures. This way of concentrating mainly on standardizing the outlet load at the basis node stage is the boundary to the present accessible procedures. It is already known that as a reason of link crash and congestion packet loss in the network routing largely occurs. In this article a new cross layer and path restoration procedure has been put forward. We also put forwarded two algorithms namely Path discovery Algorithm and congestion handling algorithm. In this approach of cross layer it comprises of 3 kinds of layers called network, MAC and transport layers. In this introduced approach the MAC and network layers have dynamic functionalities in identifying the congestion and standardization where the functionalities of network and transport layers are distinguished in bearing the congestion i.e. congestion endurance. The produced tentative results illustrate an enhanced management of congestion and its endurance by this approach.**

Keywords- manet; routing protocol; congestion control; zone; occ; cross layer.

I. INTRODUCTION

MANETs are recognized for their influence on protocols and protocol stacks of managing methods and they are naturally unsuited for customary TCP [17]. As a result ordinary TCP congestion management that is applies for the internet is not apposite method. In MANETs the nodes move fundamentally through a different means that is not known which generally is an effect of communal wireless multi hop channel and it does not get interpreted as the congestion is vanished. As a result the packet delivery delay and crash takes place. "An individual node can be transferred in its intrusion series" is the principal policy of the wireless multi hop channel. Coming to the MANETs network the whole region in the medium is packed and congested as it is common region, but coming to the internet the crowd is on main Pathr [17]. The significant character of the MANETs is that the region may be packed full but not the nodes [17].

The contrast between ordinary TCP and MANETs is due to verity that crash and losing of packets is constantly not due to the congestion in the network and the transfer periods (along with round trip periods) diversify creating complexity in identifying the lost packets. Distinct consumer is able to turn out a congestion ensuing in relatively lesser bandwidth of mobile ad-hoc networks, due to which it is complicated origin of congestion in a multi hop network. A reasonable congestion control scheme should be in use effectively to be firm and for greater functionality [17] of the wireless system as a reason of vulnerability towards congestion troubles in contrast with conventional wire line systems.

The multi-hop wireless networks are unable to attain a distinct and integrated procedure for the trouble in the congestion as they are heterogeneous character of function protocols. As an alternative an appropriate model of congestion control should be intended that mainly concentrates on the characteristics that are associated with the network [17]. Eventually these schemes serve as division of the procedures for the troubles that are recognized instead of an absolute immediately utilized method. They create themselves as the head of the customized application protocol stacks. But there are only fewer characters that work with a huge series of functions [17] and these are rare cases.

Congestion managing techniques focused mainly on the modeling, analysis, algorithm improvement of clogged loop control formats (e.g. TCP) have been observed in the latest times preparing them to get customized to the mobile hoc networks by stipulating constraints concerned to routing path and bandwidth algorithms that hold the capability to unite and improve functions are developed. There is one more chief part that should be taken into the view in the field of wireless hoc network is by reason of the MAC [Media access Control] layer [17]. That condition is that maximum wireless MACs consist a time restraint which allows the customer to utilize the physical medium but in the provided time period.

This article is structured in the method as given: The section II looks at the mainly quoted works in the field of literature. The conversations of the projected procedures are featured in section III and section IV exposes the replications and the corresponding outputs that are tagged by end conclusion and references.

II. Related Work

The mechanism for QOS centric congestion management is present in [1]. Et al, [2] initiated metrics to evaluate data-rate, MAC overhead along with buffer interruption, that assists in recognizing and manage the congestion conflicted region in network system. Hongqiang Zhai, et al, [3] put forwarded a procedure opposing that congestion and strict means conflict is interconnected. Metrics based solution on congestion aware routing was introduced in [4]. Hop stage congestion managing design was projected by Yung Yi et al, [4]. Tom Goff, Nael et al, [5] conferred a group of algorithms that instigates substitute pathways utilization in case of uncertainty in the value of a pathway that is being utilized. Xuyang et al, [6] obtained a cross-layer hop-by-hop congestion control method which was modeled for the TCP functioning in multi-hop wireless networks. Dzmitry et al [7] introduces the crash congestion at the transport layer which minimizes the functioning of the system. Duc et al [8] has discovered that the models that exist in present days cannot be adjustable for the congestion.

Discovering the congestion clearance in routing pathway is the main goal of present day methodologies. The packet crash is the cause for the link collapse. The efficient way is to work on the method to manage the packet losing which is the reason for the link collapse. Standardizing the way out near every node involving in routing is one more costly methodology. The management of congestion is near hop stage [4] [15] in most of the situations. As a result the standardizing way out method at every node in the network includes the usage of the costly possessions. This article shows the effectiveness in discovering the crash of the packets which are recurring as a reason of conflict or by buffer filled up or by malevolent fail. So the method of managing the congestion through standardizing the way out can be evaded in the conditions like the link collapse and in bitter situations this is handled by using substitute pathway restitution. We can also urge that a situation may arise where the hop stages cannot standardize themselves for which only hop stage congestion management is not enough. The way of managing the outlet force that is being followed in origin stage standardization design, can also be followed as well in the management of the congestion by using the similar possessions.

In this article introduction to a new cross layer congestion control design has been made which involves:

- The node capability and possession's heterogeneity.
- Congestion related packet crash being confirmed by cross layer design.

III. OCC: Ordered Congestion Control Using Cross Layer Support In Manet Routing

We know that in MANETs crashing of the packets happen frequently. The main causes for this to happen are as follows:

- Link collapse during transfer.
- Minimizing the packet entrance power by utilizing conditional Transfer with overwhelmed Ingress. This is also named as packet sinking because of congestion near routing.
- Medium usability conflict.
- Malevolent sinking near the recipient.

A concise explanation on introduced OCC is as given: The congestion control methodology that was put forward is attained in stratified way.

In our methodology, at first reduce channel current near the pathway node pn_p antecedent to pathway node pn_c that is affected by congestion. This step is voluntary and probable delay threshold at pn_p and functional part of buffer capability. If there is any situation of error or crash in the functioning of the primary step, then automatically this gives rise to the functioning of the secondary step of the methodology. Coming to the secondary step the MAC layer makes the adjacent nodes pn_p attentive that are also present in that particular region. As a result the outcome of all the other adjacent nodes pn_p will be reduced at a time, so that there will be no delay in the group of threshold value.

Even if the affected node has not improved after the commencement of the first steps of the methodology then the thirds step gets instigated. The procedure in this step is that the MAC checks the inward rush of the nodes I near the particular pathway pn_p in a given period of the time span τ , then the nodes that are present in that particular zone c_c of the routing path will be intimated about the affected node pn_v by the MAC. Now all the rest of the nodes reduce their outward rush in order to make the delay of the threshold group gets decreased. When the MAC checks the inward rush of the nodes I near the particular pathway pn_p in a given period of the time span τ , if $I' \geq I$ and the affected node is not improved then the pathway is re-established by making a link among the nodes pn_p and rpn_c , where rpn_c is the pathway node that is held back, which is a consideration node for pn_c . As a result the routing information avoids the affected node pn_v , that is pn_c .

A. Dividing the network in to zones

Mohammad M. Qabajeh et al [8] explained a methodology, which was chose by us. By understanding the already present nodes, the total area is separated into divisions. In general the outline of the zone is preferred as hexagon. By considering the hexagon the benefit is that the outline touches the larger area and along with it the correspondence with adjacent nodes is also easier as the outline is similar to that of the transmitter. The point-based is applicable in the MANETs due to the accessibility of reasonably cheap and portable less consumable GPS receiver. The series of the transfer for the node is represented using R and the length of the hexagon by using L. For the correspondence of the zones amongst them a relation between R and L is generated as L=R/2.

All the zones have their identification as (cid). The zones are provided with couple of alternative units. They are:

1) If a node is involved in the routing and its zone is present in the routing path, if that node has entered into the given series region then that from then is referred as pathway node.

2) If the node rpn of the pathway node pn is recognized as the substitute node for the node pn by the transport layer then that node is named as the reversed pathway node.

If the placement of the node is found, that is at which point it is present then, nodes can execute our self-mapping algorithm of the substantial site on present zone and compute the cid with no trouble. Figure 1.shows the Common outline of the Zone divisions in network.

Pathway nodes and their respective reserved pathway nodes are present for every zone.

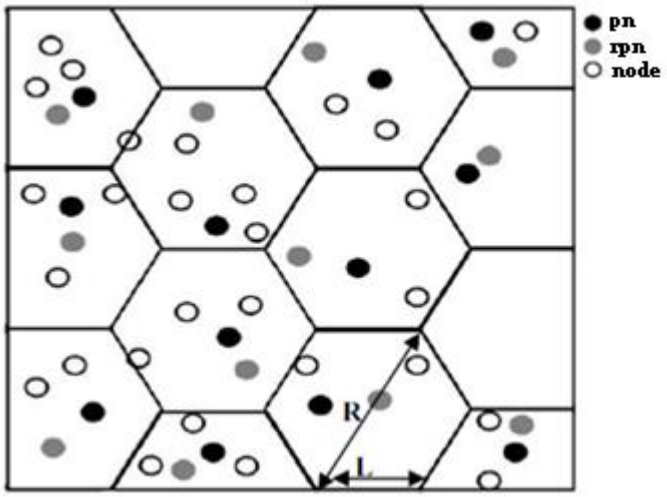


Figure 1. Common outline of the Zone divisions in network[8]

B. Path Detection

This practical methodology is termed as DSR policy for Path detection. A distributed technique is utilized in order to determine the path to the end node n_d by the source node n_s . The appealed packet $rreq$ that is being transmitted will take the node related data like the involvement in the routing path and its id value cid of that node that is communicating. While the packets are transferring the, transport layer checks the zone stage nodes of every node that is communicating and holds the data with the packet $rreq$ .After the final end node gets this packet $rreq$ from the origin then it gets ready to send the reply packet $rrep$ which includes the record of all the pathway nodes and their communicating nodes in the area of the zone. At the time when the reply packet is acknowledged than all the communicating nodes make the necessary changes in their routing table and revise it with the antecedent and descendant node data. It also revises with the other communicating nodes of that particular node and its descendant node in the pathway of the pathway.

When the reply packet $rrep$ finally reaches the origin node n_s , then the most desired path will be chosen. Then the origin node n_s delivers an acknowledgement $ack(pn)_i$ for every path node for the routing desired path. After the acknowledgement packet $ack(pn)_i$ is delivered then ahead the pathway node pn_i determines the desirable paths among the node pn_i and the both hop stage descendant node pn_{i+2} . In this step the main path node pn_i delivers an appeal $rreq$ to pn_{i+2} . This appeal $rreq$ communicates by using only the communicating nodes of the main node pn_i and the node pn_{i+1} . When this appeal is delivered to pn_i , then pn_{i+2} acknowledges it by using the packet $rrep$ and transfers it to the pn_i on the same path that used by the $rreq$. When the acknowledgement $rrep$ is delivered then pn_i chooses the desired path among the nodes pn_i and pn_{i+2} , lastly accumulating it into the routing tables. The desired path that was chosen is utilized for the re-establishment among the nodes $pn_i \; and \; pn_{i+2}$, on the basis of a condition that the congestion is obvious at the adjacent descendant node pn_{i+1} of the main node pn_i .

Path detection algorithm

1. n_s Creates $rreq$ and transmit it to adjacent units.
2. When $rreq_i$ is delivered, hop stage node n_i confirms that whether retransmitting of $rreq_i$ is previously completed on their own or not.
3. If retransmitting is previously completed then rejects the $rreq_i$, or else n_i gathers the particulars of communicating nodes from transport layer and along with that it includes its own recognition and particulars of its communicating nodes to $rreq_i$, then retransmits . This procedure continues until $rreq$ is delivered to the end node n_d
4. Then end node n_d creates acknowledgement packet $rrep_i$ that includes the particulars of the nodes that are present in the pathway. By utilizing this acknowledgement $rreq_i$ navigated to arrive at n_d and it's communicating nodes. The acknowledgement

packet $rrep_i$ transfers in reverse to the origin node n_s on the same desired path of the packet $rreq_i$.

5. Every transitional node pn_i in the path which utilized the packet $rrep_i$ gathers the particulars of its antecedent node pn_{i-1} in pathway, descendant node pn_{i+1} and communicating nodes of main path node pn_i and descendant path node pn_{i+1}
6. Main path node pn_i revises its routing table by the particulars attained in earlier step.
7. The methods 6 and 7 frequently continue until acknowledgement packet is delivered to origin node n_s
8. Origin node n_s derives the desired path which includes zones with crowded nodes.
9. For every pathway node 1 to n of the path chosen, n_s replies $ack(pn)_i$ for $i = 1..n$.
10. On gaining the $ack(pn)_i$, pn_i begins deriving the substitute path among pn_i and pn_{i+2} , so that the substitute path can utilize communicating nodes of the pn_i with pn_{i+1} merely.
11. pn_i then records substitute path among the nodes pn_i *and* pn_{i+2} at routing collection.

C. Managing Congestion

When the packet is crashed and that is determined that it is crashed at the node pn_i then MAC layer checks the conflict position near pn_i , if that point is found then it makes the antecedent node pn_{i-1} of main node pn_i aware regarding the need of the retransferring in the given span τ as conflict alert con_+ . If the span id maximizing the delay near pn_{i-1} so that the packet is crashing at the node pn_{i-1} and its values is higher than the tolerable threshold value, then the node depends on the substitute path among the nodes pn_{i-1} and pn_{i+1} which is present in the routing collection. This substitute path will be on use until the MAC layer sends the acknowledgement con_- of conflict removed at the main path node pn_i by the node pn_{n-1} . When the node pn_{i-1} receives the acknowledgement mac_- sent by the MAC layer then it returns the path back to the pn_{i+1} . The MAC layer again validates and if it derives that the congestion is not because of the conflict then MAC checks the buffer during the inward rush at the main node and if it is full then delivers bof_+ regarding the crowd in the buffer. When the node pn_{i-1} takes delivery of bof_+ , then it tries to reduce the inward rush so that the delay that is incrementing may not make the packet get crashed at the main node pn_{i-1} . If this process fails to reduce the inward rush at the node pn_{i-1} , then the network layer makes the all the remaining nodes of the zone c_c in which the node pn_{i-1} is present aware, so that the rest will reduce their inward rush because of which the increment in the delay may not make the packet get crashed near their own man nodes. Even if this case fails than the network layer makes the descendant zone of the present zone c_c aware of this situation. This procedure will be continued frequently until the congestion that is caused due to the rush in the buffer gets prohibited or it is delivered to the zone c_s in which origin node n_s is present. If the result is failed to come then in order to continue the information transfer among the nodes pn_{i-1} and pn_{i+1} , that was troubled because of the congestion bear the main node pn_i , the node pn_{i-1} depends on the substitute path that is accessible in the routing collection. If the MAC determines that the congestion is occurred due to the link collapse among the node n_i and its descendant n_{i+1} then the main node pn_i chooses the substitute path in order bond up with the pn_{i+2} that is stored in the routing collection of the main node pn_i .

Algorithm for congestion management and working information transport in opposition to congestion

1. Let us assume a case of packet crashing at pn_i
2. MAC checks the position of the conflict:
3. If congestion arouse because of conflict near pn_i
   a. then MAC recognizes the conflict near pn_i and make pn_{i-1} aware by sending a information in con_+ ,
   b. then pn_{i-1} functions on the congestion caused by the conflict: move to step 6.
4. else if congestion is caused because of the rush in buffer near pn_i
   a. then MAC recognizes the rush in buffer near pn_i and make pn_{i-1} aware by sending a information in bof_+
   b. then pn_{i-1} functions on the congestion caused by the rush in buffer: move to step 7
5. else if congestion is caused because of the link

collapse among pn_i *and* pn_{i+1} near pn_i

a. then MAC recognizes the link collapse among pn_i *and* pn_{i+1}, and make aware by sending a link collapse information in LF_+.

b. then path node pn_i functions on the congestion caused by the link collapse: move to step 8

6. Managing Congestion caused by conflict:

i. When con_+ is taken the delivery from MAC, path node pn_{i-1} functions

a. Evaluate the con_+, that includes a particulars regarding whether retransfer is needed and span τ for retransfer.

b. Checks the weight of the τ on inward rush delay time Δ

i. If $\Delta \geq \delta$ (inward rush delay threshold) [consequences are packet termination because of surpassed delay] For span τ, choosing to substitute path among path node pn_{i-1} and pn_{i+1} to avoid the affected node pn_i, which was caused by congestion by conflict.

ii. Past the span τ path node pn_{i-1} is taken the delivery of either con_+ or con_- from MAC. MAC delivers mac_+ if conflict is still id present in the affected node pn_i else intimates to pn_{i-1} regarding the situation of no conflict at affected node pn_i through con_-.

iii. If con_+ is delivered from MAC then pn_{i-1} executes steps 1 and 2.

iv. else if con_- is taken delivery by pn_{i-1} then it re-establishes the original path among pn_{i-1} *and* $pn_{i+!}$

7. Managing Congestion caused by the rush in the Buffer

v. When bof_+ is taken the delivery from MAC, path node pn_{i-1} functions

Evaluate the bof_+, that includes a particulars regarding congestion because of rush in the buffer near pn_i.

vi. Executes the procedure of inward rush reducing so that delay Δ does not cross delay threshold δ limit.

vii. If inward rush not reasonable as needed to manage the congestion near pn_i then

a. Network layer makes every path node that is located in the similar zone c_c to which pn_{i-1} is part of, aware regarding congestion position near pn_i.

b. As a result every path node of zone c_c tries to reduce their inward rush so that that delay Δ does not cross delay threshold δ limit of individual path nodes.

viii. If inward rush near individual nodes not reasonable as needed to manage the congestion near pn_i then

a. Network layer makes path nodes in the zone c_p aware, that is antecedent to the c_c.

b. As a result every path node of zone c_c tries to reduce their inward rush so that that delay Δ does not cross delay threshold δ limit of individual path nodes.

c. If $n_s \notin c_p$ then $c_p \rightarrow c_c$: move to step viii.

d. Else if inward rush at individual not reasonable as needed to manage the congestion near pn_i then pn_{i-1} chooses the substitute path that bonds pn_{n-1} *and* pn_{n+1} to make the information transport, which avoids the congestion affected node pn_i.

8. Managing congestion caused by link collapse

ix. When LF_+ is taken the delivery from MAC then path node pn_i chooses the substitute path alp that bonds n_i *and* n_{i+2} to make the information transport.

In view of the fact that the alp is being utilized the path node pn_i tries to derive a desired path among pn_i and pn_{i+2} and this substitute path gets constructed by considering communicating nodes of pn_i *and* pn_{i+1}.

IV. Simulations And Results Discussion

The tool that was utilized in accomplishing the test was NS 2. Considering the mobility and amount ranging from 20 to 200, a simulation network simulation network has been constructed. The attributes and the values of the simulation are explained in the below table 1. If the packet that was sent is legal then it confirms that buffer is assigned successfully. The main goal of this model is to contrast the congestion and contention control protocol [18] and OCC. The functionality test for the two protocols by using the metrics given as follows:

TABLE I. SIMULATION ATTRIBUTES TAKEN FOR THE TEST

Amount of nodes Range	50 to 200
Dimensions of space	1500 m × 300 m
Nominal radio range	250 m
Source–destination pairs	20
Source data pattern (each)	4 packets/second
Application data payload size	512 bytes/packet
Total application data load range	128 to 512 kbps
Raw physical link bandwidth	2 Mbps
Initial PATH REQUEST timeout	2 seconds
Maximum PATH REQUEST timeout	40 seconds
Cache size	32 Paths
Cache replacement policy	FIFO
Hash length	80 bits
certificate life time	2 sec

There are few metrics in order to examine the working of the approached methodology. They are as follows:

- DATA PACKET DELIVERY RATIO: The ratio is derived by computing the division between the amount of information packets delivered from source and the amount of the information packets acknowledged by sink.
- PACKET DELIVERY FRACTION: The ratio between the information packets that are carried to target area and the packets that were produced from the origin. This paper gives the information on the working of the methodology showing the effectiveness in carrying the packets to the target. As the rate of working increases it produces the more accurate outputs.
- AVERAGE END TO END DELAY: This is defined as the common peer and peer delay of the information packets. Some of the reasons for this problem are loading at the time of path identification, LIFO near the interface queue, resending delays storing at the MAC and transport period. By this the differentiation of the time periods when the packet was sent and reached is determined. After this time period is derived, separating the whole time period differentiated value on the overall amount of CBR packets reached delivers the end-to-end delay for the packets that were reached. As the delay decreases the functioning, the working of the methodology is enhanced when the delay period is less.
- PACKET LOSS: This ratio is derived by subtracting the amount of packets that were delivered at the origin and the amount that reached to the sink. There were failure packets in the output received by us at Network and the MAC layers. This method sends the packet to the target until the path is derived, or else it searches unless a path is figured out. There are also situations when the buffer leaves the packet, couple of them are: Packet is ready for the search of the path, but the buffer is not empty. The second one is the search of the path crossed the time limit. From this, it is derived that as much less is the loosing of the packets, more will be the functioning of the approach.
- ROUTING OVERHEAD: This is termed as the ratio of overall amount of the routing packets and the information packets. This ratio is determined at the MAC layer.

Figure 2(a) illustrates Packet Delivery Ratio (PDR) for Congestion and Contention Control Protocol [18] and OCC. By considering this output it is enough to prove that OCC manages maximum failure of PDR than that of [18] in case of DSR. Fairly accurate failure amount of PDR that is restored by the OCC than [18] is 1.5%.

This is balanced amount among the pauses. The least amount of restoring examined is 0.18% and the highest id 2.5%. The next Figure 3(b) specifies OCC benefit than that of [18] in case of Path optimality. [18] utilized nearly 0.019 hops more when compared to OCC as the reason of dual distribution of the [18].

The derivation for the packet delivery fraction (PDF) is:

$$P' = \sum_{f=1}^{e} \frac{R_f}{N_f}$$

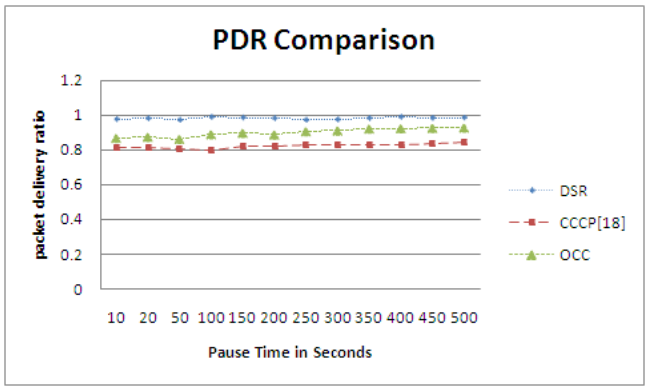
$$P = \frac{1}{c} * P'$$

- P is the division of effectively reached packets,
- c is the overall amount of flow or associations,
- f is the distinctive flow id allocated as index,
- R_f is the amount of packets acknowledged from flow f
- N_f is the amount of packets transferred to flow f

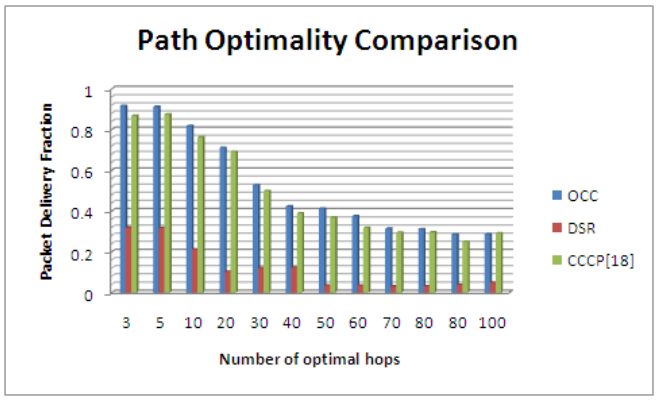
Figure 2(c) proves that OCC is has less packets than that of [18]. This benefit of the NCTS could be feasible as a reason of availability of constant paths without negotiation or offended nodes and having cross-layer congestion control and effective routing procedure. The Packet transparency derived in [18] is nearly 5.29% larger than packet transparency derived in OCC. The least and highest packet transparency in [18] than OCC derived is 3.61% and 7.29% correspondingly.

MAC load transparency is high in OCC than [18] to some extent. This is viewed in figure 2(d). This is occurred due to the control packet swap in OCC. The common MAC load transparency in OCC than [18] 1.64%. The least and highest MAC load transparency derived is 0.81 and 3.24% correspondingly.

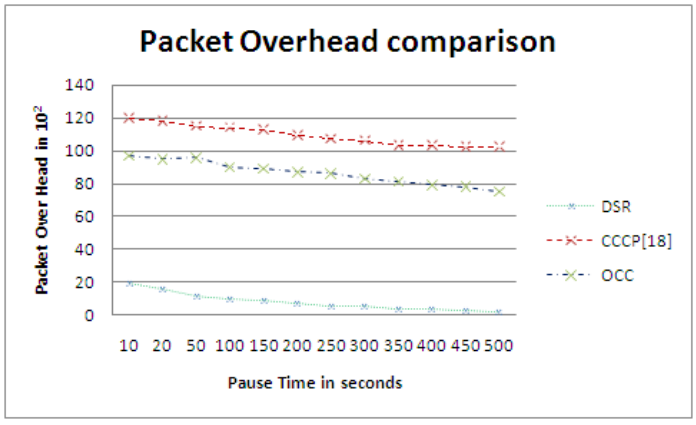
Appealing outputs have been determined for DSR when all the assessment procedures are considered. Apart from path optimality DSR executed fine as a result of not taking security concern into account as a routing attribute, and it is producing enhanced QOS without risk in routing hypothesis. But factually it is false in actual. In path optimality verification DSR place at end as a reason of not taking security restraints into account, amongst three measured procedures, eventually this made to recognize uneven paths.



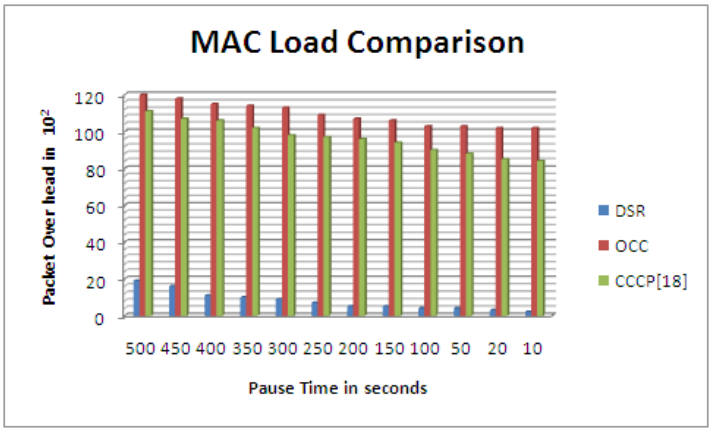
(a) Packet delivery ratio assessment utilizing line chart



(b) Bar chart illustration for Path optimality



(c) A line chart illustration of Packet transparency assessment details



(d) Mac load assessment illustrated in bar chart layout

Figure 2. Assessment details for OCC functioning than CCCP [18]

V. CONCLUSION

This article talked about the routing algorithm named as "Stratified cross layer congestion control and endurance routing protocol". This proposed procedure derivate two algorithms for Path discovery and congestion management correspondingly. Conventional proactive routing protocol DSR is utilized to derivate Path discovery algorithm. The congestion managing procedure has been separated into three units. Primary one manages the congestion occurred because of the conflicts, secondary unit is to manage the congestion occurring because of rush in the buffer and the final unit is to manage the congestion because of link collapse.

In the type of the congestion because of the rush in the buffer, our procedure manages it near the antecedent path node stage and error in this situation solves it by considering first to the zone stage and next to the network stage. This chronological procedure reduces the effort and cost for the consumption. The path recovering at the node stage that was chosen in managing the congestion facilitated the possibility of information transport opposing the rigorous congestion caused because of conflict and link collapse.

The experimental outputs that were obtained were efficient and noteworthy, such that we can widen the boundaries of the application in reducing the delay and improving the cross layer mechanism for effortless competence in the future course of time.

REFERENCES

[1] [1] Michael Gerharz, Christian de Waal, and Matthias Frank, “A Practical View on Quality-of-Service Support in Wireless Ad Hoc Networks”, BMBF

[2] Xiaoqin Chen, Haley M. Jones, A .D .S. Jayalath, “Congestion-Aware Routing Protocol for Mobile Ad Hoc Networks”, IEEE, 2007

[3] Hongqiang Zhai, Xiang Chen, and Yuguang Fang, “Improving Transport Layer Performance in Multihop Ad Hoc Networks by Exploiting MAC Layer Information”, IEEE, 2007

[4] Yung Yi, and Sanjay Shakkottai, “Hop-by-Hop Congestion Control Over a Wireless Multi-Hop Network”, IEEE, 2007

[5] Tom Goff, Nael B. Abu-Ghazaleh, Dhananjay S. Phatak and Ridvan Kahvecioglu, “Preemptive Routing in Ad Hoc Networks”, ACM, 2001

[6] Xuyang Wang and Dmitri Perkins, “Cross-layer Hop-byhop Congestion Control in Mobile Ad Hoc Networks”, IEEE, 2008

[7] Dzmitry Kliazovich, Fabrizio Granelli, “Cross-layer Congestion Control in Ad hoc Wireless Networks,” Elsevier, 2005

[8] Duc A. Tran and Harish Raghavendra, “Congestion Adaptive Routing in Mobile Ad Hoc Networks”, 2006

[9] Nishant Gupta, Samir R. Das. Energy-Aware On-Demand Routing for Mobile Ad Hoc Networks, OPNET Technologies, Inc. 7255 Woodmont Avenue Bethesda, MD 20814 U.S.A., Computer Science Department SUNY at Stony Brook Stony Brook, NY 11794-4400 U.S.A.

[10] Laura, Energy Consumption Model for performance analysis of routing protocols in MANET,Journal of mobile networks and application 2000.

[11] LIXin MIAO Jian –song, A new traffic allocation algorithm in AD hoc networks, “The Journal of ChinaUniversity of Post and Telecommunication”, Volume 13. Issue3. September 2006.

[12] Chun-Yuan Chiu; Wu, E.H.-K.; Gen-Huey Chen; "A Reliable and Efficient MAC Layer Broadcast Protocol for Mobile Ad Hoc Networks," Vehicular Technology, IEEE Transactions on , vol.56, no.4, pp.2296-2305, July 2007

[13] Giovanidis, A. Stanczak, S., Fraunhofer Inst. for Telecommun., Heinrich Hertz Inst., Berlin, Germany This paper appears in: 7th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, 2009. WiOPT 2009

[14] Outay, F.; Vèque, V.; Bouallègue, R.; Inst. of Fundamental Electron., Univ. Paris-Sud 11, Orsay, France This paper appears in: 2010 IEEE 29th International Performance Computing and Communications Conference (IPCCC)

[15] [Yingqun Yu; Giannakis, G.B.; , "Cross-layer congestion and contention control for wireless ad hoc networks," Wireless Communications, IEEE Transactions on , vol.7, no.1, pp.37-42, Jan. 2008

[16] http://www-lih.univ-lehavre.fr/~hogie/madhoc/

[17] Prof.K.Srinivas and Prof.A.A.Chari. Article: Cross Layer Congestion Control in MANETs and Current State of Art. International Journal of Computer Applications 29(6):28-35, September 2011. Published by Foundation of Computer Science, New York, USA

[18] Yingqun Yu; Giannakis, G.B.; , "Cross-layer congestion and contention control for wireless ad hoc networks," Wireless Communications, IEEE Transactions on , vol.7, no.1, pp.37-42, Jan. 2008 doi: 10.1109/TWC.2008.060514.

Multi-Objective Intelligent Manufacturing System for Multi Machine Scheduling

Sunita Bansal
Research Scholar
JJTU University,
Rajasthan, India

Dr. Manuj Darbari
Associate Professor
BBD University,
Lucknow, India

***Abstract*—This paper proposes a framework for Intelligent Manufacturing systems in which the machine scheduling is achieved by MCDM and DRSA. The relationship between perception/knowledge base and profit maximization is being extended. Further for production function.**

Keywords-Decision making; pareto; intelligent manufacturing.

I. INTRODUCTION

The Multiobjective function has the characteristic that there exists no fixed solution to defined problem instead it could have multiple solutions. The correlation between objectives is a very complex phenomenon and depends on the alternates available.

The multiobjective optimization consists of three phases: Model Building, optimization and decision making. To solve the problem following the true multi-objective functionality interface optimization and preference management.

In this paper we will be discussing on the issues relating to Honda Car manufacturer with Multi Assembly Line. The focus is to optimize the part versus product planning. Firstly we develop the part description which starts from the component level manufacturing.

II. LITERATURE SURVEY

Previous studies in the field of Decision Maker (DM) have shown good solutions in a given problems. There has been good research by Pinedo and chao (1999) in the field of flexible assembly system. Cochram et al (2001) has also analyzed how the selection of manufacturing system.

Further Seward and Nachlas (2004) also considered availability in the analysis of manufacturing systems[2]. A searching and sorting choice was been analyzed previously by Jasz kiewicz and Ferhat (1999), they later on modified it for Multiobjective[1] optimisation method. later on posterior rationality in MCDM was presented by Greco, et al (2008).

III. IMPLEMENTATION OF MULTI -OBJECTIVE FUNCTIONALITY IN INTELLIGENT MANUFACTURING

The Main focus in intelligent manufacturing system is to represent the parts in terms of its features. The feature deals with all the geometric and technological information of each feature.

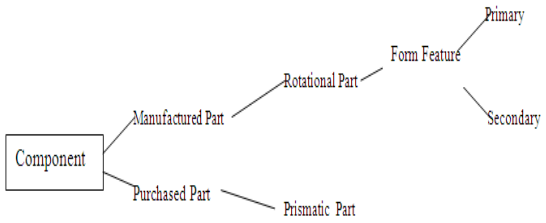


Figure 1 : Part definition object Hierarchy (Adopted from OO Approach to Features based process Planning, John H. Usher)

We handle the above part definition by applying Multi-Objective optimization as:

Minimize {f1 (Primary features), f2 (Secondary features)} (1)

is approach is solved by the help of Pareto optimal solution. To proceed further we define the following steps:-

1) Initialize
2) Generation of Pareto Optimal Starting Point
3) Decision making preference is sorted out.
4) Generation of New Pareto optimal Solution according to form features.
5) If several possible solutions are generated than signal the Decision Making to stop.
6) STOP

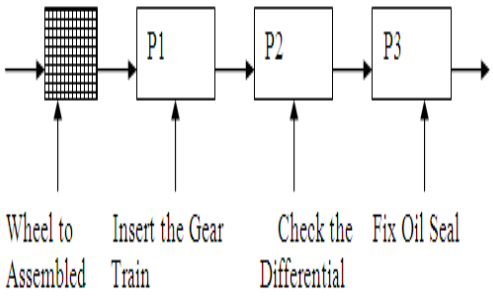


Figure 2. Assembly Line for Wheel Base

Consider layout of Toyota Fabrics

Unit where final wheel assembly has to be done. The possible operations that can be done[3]. The possible operations that can be done to assemble the wheel deals in the following manners:

Wheel Assembly[ ] [ ] = {{ "Insert Gear, “Align Train" Gear”},{"Insert" Check Cross"; Alignment"}, {"Fill Oil Check to the specified Level”Oil Seal}} (2)

Within the process hierarchy each operation can be treated as an entity which expresses the operation. These operations are solved by the use of Interactive Multiobjective optimisation using Pareto principles as stated previously. In the first stage Pareto optimal set is generated from the Machine tool settings. In the second stage, the Decision Makers (DM) provides the necessary solution to the above problem as shown in figure 3. The Heuristic solver uses" if - then " decision rules. These rules

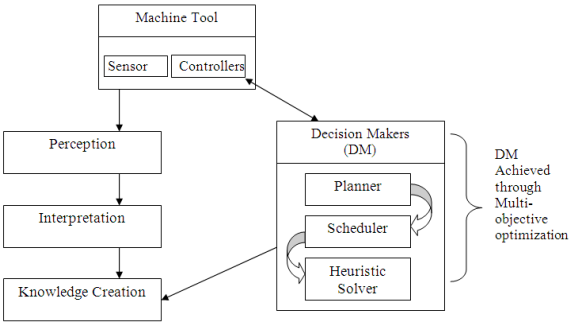


Figure 3 : Machine Tool and Decision Maker connectivity

Provide a solution from the currently considered optimal set.

It is very difficult to prioritize all of the objectives or set of goals. We use relational operator which uses the goals and priority information[4]. The ranking of whole population is based on a relationship theory.

A. Implementation

Consider an n- dimensional vector defined for wheel Assembly [ ] [ ]. With decision variable Maximize Output and n- dimensional objective vectors Perception = f (Maximise output) and Knowledge Creation = f (Maximise Output) are the two extreme requirements of Maximisation of output. Consider a preference Vector of Heuristic set rules defined as:

h = {h1.,...hp}= { (h11,... hn)... (hp1.....hpnp)} (3)

Now applying this principle of heuristic to "Perception" and "Knowledge Creation" as:

hpk = { hp11, ...hpn1)..(hk11,..... hpkp)} (4)

The sub vector 'hp' of the associate vector represents the priorities from 1,..... p with goal f (Maximise Output)[5].

Hence a new Pareto optimality[ ] for Convexity is obtained as:

!perception, knowledge Creation ∈ Maximise Profit.

$$\left[F_M(x) = \max_{i \in (1...n)} \left\{ \frac{f_i(x) - (Maximise\,\Pr ofit)_i}{(Perception \wedge Knoweldgecreation)_i} \right\} \right] \quad (5)$$

From Equation 5 we can visualize the trade off between Profit Maximization and Perception, Knowledge creation as shown in figure 4.0

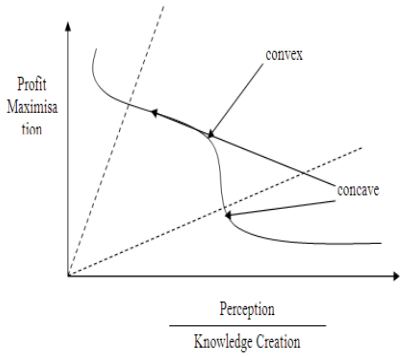


Figure 4 : Trade-off between Profit Maximization and Perception and Knowledge Creation.

The factor of knowledge creation and perception is derived from Decision maker which has a component of Heuristic solver.

Heuristic solver user various set of rules with following variables :

$$Y = f\{L, K, R, S, \nu, \gamma\} \quad (6)$$

Where Y = Output
L = labour Input
K = Capital Input
R = Raw Materials
S = Land Input
ν = Return to Scale
γ = Efficiency parameter

The heuristic model us Reference Point Substantial modeling technique which decides on the trade off in the factors described in Equation 6, it is given as:

$$y = F(\nu, \gamma)\ ;\ x \in S \quad (7)$$

where

Y = vector of outcomes, used for measuring the consequences of implementation of decisions.

x = Vector of decisions like L, K, R and S. which can be controlled by the user.

γ & ν = Vector of external impact which is not under users control.

F = Vector of functions like objective & constraints.

S = Set of feasible decisions.

To analyse this the vector outcomes the condition attributes are criteria and decision classes are preference ordered which means this vector of decisions like x, γ and ν moves to upward and downwards unions of classes[6].

According to DRSA if A dominate B with respect to PCI it is denoted as :

ADPB for every criteria

$$i \in P, fi(A) \geq fi(B) \quad (8)$$

The granules of knowledge used for approximate are: a set of objects dominating A, called P - dominating set

$$D+p(A) = \{B \in U : BDPA\} \quad (9)$$

A set of objects dominated by A, called P-dominating set

$$D+p(A) = \{B \in U : ADPB\} \quad (10)$$

Now let A = {x}; B={γ, ν} and Cl = { Cl1…Clm}

denotes decision classes sorting such that each A ∈ U belongs to one and only one class Ct.

"≥" is a comprehensive weak preference on U if for all A, B ∈ U, A ≥ B reads as " A is at least as good as B" which means

[A ∈ Clr, B ∈ Cls r >s]yA > B]

Where Clr are preferred to the object from Cls For every P ⊆ I, the quality of approximation at sorting Cl by a set of criteria "P" is defined as The ratio of the member of object P - consistent with the dominance principle and the number of all the objects in U[7].

The quality of approximation of sorting Cl with criteria as maximisation of Profit[8].

There are certain decision classes which we will derive in terms of "if .... then" (a part of Heuristic Solver). According to DRSA a given upward or downward union of classes Clt ≥ or Cls≤ the decision rules induced a hypothesis that objects belong to P (Cl t ≥) or P (Cls ≥)[9].

There are three types of decision rules :

1) Certain D≥ - decision rules, providing lower profile of the objects like γ and ν .

2) Certain D<-decision rules providing the maximum values of the considered criteria like L,K,R and S.

3) Approximate D≥ ≤- decision rules, providing simultaneously lower and upper profiles of objects like x versus γ and ν.

A set of decision rules is complete if it represents all the objects[10].

Applying these rules to "Theory of Production". We have the plot between the two dominated objects as : x and γ and ν.

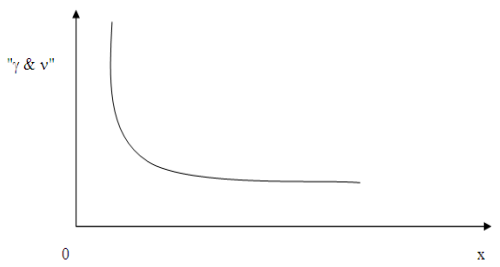


Figure 5 : Under Normal condition the relationship between two objects "x" and " γ & ν".

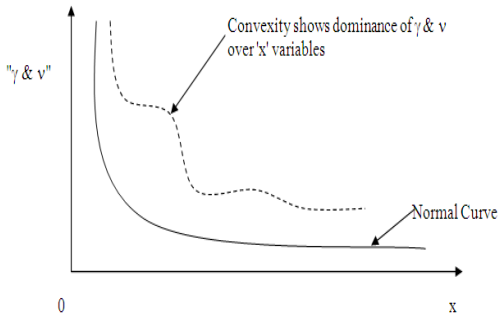


Figure : 6 Optimised graphical Plot showing the Hypothesis

The convexity shows dominance of optimizing "return to scale" and "Efficiency parameter".

IV. Conclusion

The paper gives a brief overview of the Pareto optimal set and Decision maker (DM) which shows a good solution to the production problem in which we have DRSA to find out the dominance of the variables in the factor of production.

By solving using the dummy variables we are able to derive the optimised graph of the solution vector for production function.

References

[1] Aguirre, H. E., Tanaka, K., Sugimura, T. and Oshita, S. (2001). Halftone image generation with improved multiobjective genetic algorithm. In Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization (EMO-2001), pp. 510-515

[2] Antonisse, J. (1989). A new interpretion of scheme notation that overturns the binary encoding constraint. In Proceedings of the Third International Conference on Genetic Algorithms, pp. 86-91.

[3] Arora, J.S. (1989). Introduction to Optimum design, New York: McGraw-Hill

[4] Bagchi, T. (1999). Multiobjective Scheduling by Genetic Algorithms, Boston: Kluwer Academic Publishers

[5] Deb, K.(1999a). Evolutionary algorithms for multi-criteria optimization in engineering design. In K. Miettinen, P. Neittaanmaki, M. M. Makela and J. Periaux (Eds), Evolutionary Algorithms in Engineering and Computer Science, pp. 135-161. Chichester, UK:Wiley

[6] Deb, K. and Goel, T. (2001b). A hybrid multi-objective evolutionary approach to engineering shape design. In proceedings of the Third International Conference on Genetic Algorithms, pp. 42-50

[7] Deb, K. . Multi-objective evolutionary algorithms: Introducing bias among Pareto-Optimal solutions. In A. Ghosh and S. Tsutsui (Eds), Theory and Applications of Evolutionary Computation: Recent Trends. London: Springer-Verlag

[8] Fonseca, C. M. and Fleming, P.J. (1998a). Multiobjective optimization and multiple constraint handling with evolutionary algorithms – Part I: A unified formulation. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans 28(1), 26-37

[9] Goldberg, D.E. (1989). Genetic Algorithms for Search, Optimization and Machine Learning. Reading, MA: Addison-Wesley

[10] Horn, J., Nafploitis, N and Goldberg, D. (1994). A niched Pareto genetic algorithm for multi-objective optimization. In Proceedings of the first IEEE Conference on Evolutionary Computation, pp. 82-87

AUTHORS PROFILE

Sunita Bansal is Research Scholar in JJT University. Her teaching areas are Artificial Intelligence, Advanced software engineering, Distributed Systems, Object Oriented Systems, SPM. She has published three international journals.

Manuj Darbari is currently working as an Associate Professor in Information Technology at Babu Banarasi Das National Institute of Technology & Management, Lucknow. His teaching areas are ERP, MIS, Soft Computing. He has published ten papers in referred international and national journals. He was selected for Marquis who's who in science and engineering 2003-2007. His teaching areas are Information Science, ERP, Software Engineering, and Workflow Management.

Evaluation of Data Security Measures in a Network Environment Towards Developing Cooperate Data Security Guidelines

Ayub Hussein Shirandula
Dr. G. Wanyembi
Mr. Maina karume
Masinde Muliro University of Science and Technology

***Abstract*— Data security in a networked environment is a topic that has become significant in organizations. As companies and organizations rely more on technology to run their businesses, connecting system to each other in different departments for efficiency data security is the concern for administrators. This research assessed the data security measures put in place at Mumias Sugar Company and the effort it was using to protect its data. The researcher also highlighted major security issues that were significantly impacting the operations of Mumias Sugar Company. The researcher used the case study methods where both qualitative and quantitative data was collected by use of questionnaire, interviewing and observation. From the findings the researcher developed data security guidelines for Mumias Sugar Company. The information gained from extensive literature review was tested and observed during the case study.The research revealed that data security lapses in the company was as a result of system administrators' failure to update and train computer users in the company on how to implement different data security measures that were in place. The final outcome of the research was data security guidelines that were practical enough to be used at Mumias Sugar Company.**

Keywords- Data, security; security measures; guidelines; computer users; Mumias Sugar Company.

I. INTRODUCTION

Today, most companies need information systems to prosper and survive for too long. Data has now become a valuable asset to modern organizations. Therefore it was imperative for Mumias Sugar Company to take the protection of their information resources seriously. (Geer, 2003), Lack of knowledge in the organization is the greatest threats to data security (Mitnick & Simon 2002). Without adequate level of user Corporation and knowledge, many security techniques are liable to be misused or misinterpreted by users which may result in an adequate security measures becoming inadequate (Sponen, 2001).

Most people believe that it is impossible to set up a computer system in a network environment and ensure that the system is secured. According to (Connolly, 2000), "the only secure system is one that is completely disconnected from a network and lying off the bottom of the ocean" it is a fact that "data security is an information technology hot issue as long as a computer is networked" (Connolly, 2000)

A. General Objective

The general objective of this research was to establish how the companies' data security could be determined by the data security measures that it had put in place and how it required different controls. In order to improve the data security of MSC, it was important to review data security measures that the company had put in place.

B. Specific Objectives

Main objectives of this research were:

1) To asses and analyze the security measures being used in the company.

2) To evaluate the security measures being used in the company with existing security standards such as ISO 17799.

3) To develop cooperate data security guidelines for the company on how they can use different security measures to ensure the adequacy of its data security.

C. Research questions

The following research questions were selected to address the stated problems.

1) Which data security measures does Mumias Sugar Company have in place?

2) How is Mumias Sugar Company using the different data security measures it has set up?

3) How can Mumias Sugar Company develop data security guidelines to use different security measures it has established?

D. Purpose of the Study

For Mumias Sugar Company to accomplish its strategy it had heavily invested in computerization, several of its departments were computerized. This research was essential for the networked environment of the company. The challenges of computerization had adversely affected the farmers and the management, where they had complained over the distortion of data in the company. Some farmers had supplied canes without payments, a fact that leads to slow the advancement of Mumias

Sugar Company. The researcher was assessing the security measures that were being used at the company.

This research contributes literature on how Mumias Sugar Company data is assured of its confidentiality, integrity and availability. A comprehensive framework was developed for Mumias Sugar Company to be able to safeguard its data and information resources from theft, abuse, misuse and any form of damage. Responsibility and accountability of information was assured at the end of the research.

E. Conceptual frame work

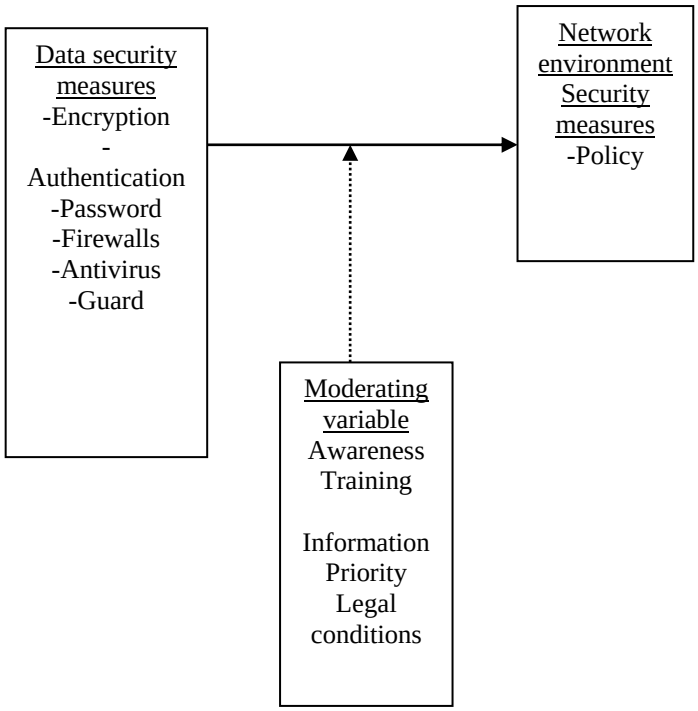


Figure 1. data security measures and its moderating variables

In the network environment existing data security measures were available. This data source was developed independently and was divided into three parts (Figure 1). Considering the complexity of securing data in a networked environment of MSC, the researcher proposed a newly designed data security measures system which was particularly designed to address the needs of the Mumias Sugar Company. By extracting the data from the original data sources the researcher improved and simplified the data security measures structure, address awareness, training, information priority and legal conditions which was important for data encryption, authentication, password, firewalls antivirus and guard (Figure 1). The conceptual framework manages the loading and updating of the data security measures.

F. Literature review

Large number of organization focuses on the technical defenses such as, encryption, access control, firewalls and intrusion detection that is associated with information protection (Ande,1972 & sand, 1996) however there was little comprehensive research that focused on how companies should: prepare for facing security incidents by selecting appropriate security measures, evaluate their present vulnerability (risk) to security incidents, asses the damages of past security incidents, train security personnel in law enforcement agencies to better prepare for dealing with security incidents.

The major problem associated with information security is that the damage is invisible and its existence is unknown. This causes difficulties for managers to justify their investments on security. (Butl, 2002 & Cohe, 1991).

The most serious financial losses in organization are related to theft of proprietary information. Information is a key resource in global competition. Organizations spent a reported 15% of the IT 2006 budget on information security and increased the rate of security staff hiring, but did not realize improvements in enterprise security, according to 2006 annual security survey (Berinato,2007).

From the literature new threats to information systems occur from unexpected sources when organizations become more reliant on it (Nyanchama, 2005). "Threat is an indication of impending danger or harm" (Johnson, 2008). "A security threat is a condition of vulnerability that may lead to an information security being compromised." (Kumar, Park, and Subramaniam, 2008).

In 2008 95% of crimes in the sphere of information were personal data thefts with 93% in 2007. Today practically everyone admits that the only way to provide information security is to take a complex approach combining measures at four levels: legislative, administrative, procedural, programmed-technical and economic.

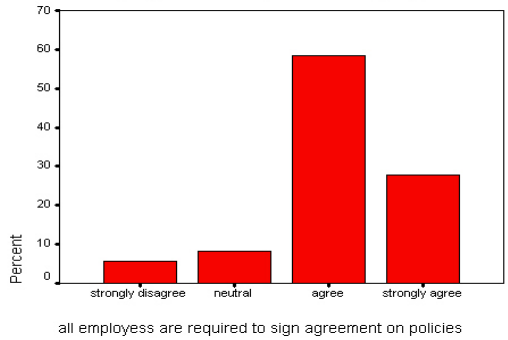
G. Qualitative and Quantitative Research Methods

Studies require different methods depending on the purpose and type of the research. Various literatures have extensively discussed the many different approaches of gathering data but most distinction is made between qualitative and quantitative methods. The aim of qualitative research methods was to explore reality through an investigation of human aspects such as feelings and behavior. In this study human behavior was observed.

While the quantitative approach explored numerical and quantifiable aspects as per (Maylor & Blackmon, 2005). In order to draw general conclusion of this research we needed measurable data (quantitative approach). But in order to increase understanding of security issues in the company in depth we needed to use qualitative approach to answer "why" and "how" questions as per (Johnson, 2004). For this research, the facts clearly suggested that we used the two approaches.

H. Views on Security Policy

1) Majority of the employees (90%) confirmed presence of a formal Policy in their company. Regarding update all information security policies were reviewed at least once a year and updated as needed. 86% of those who were interviewed confirmed that all employees were required to sign an agreement verifying they had read and understood the security policies and procedures (Refer to Graph.1)



Graph –1 All employees are required to sign agreement on policies

2) A periodic revision of the Policy brings many benefits for the company. When asked how often they revised their Policy? 73.3% interviewed replied once annually basis. However, they should be doing it on as needed basis.

3) In case of any security incident in the company response plan was formally documented and disseminated to the appropriate responsible parties. All security incidents were reported to the person responsible for security investigation. There was an incident response team ready to be deployed in case of any a data compromise.

I. Opinion regarding the implementation of control measures.

All users are required to authenticate using, at a minimum, a unique username and password which is changed after every three month. 50% strongly agreed that they change their password, however they reported that they were not trained on how to generate those passwords hence they faced problems of forgetting from time to time. (See table 1)

TABLE I. ALL USERS ARE REQUIRED TO AUTHENTICATE USING PASSWORD

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	neutral	3	8.3	8.3	8.3
	agree	15	41.7	41.7	50.0
	strongly agree	18	50.0	50.0	100.0
	Total	36	100.0	100.0	

All passwords on network devices and systems were supposed to be encrypted. From the literature review it has been observed that password used mostly is not. When asked whether all passwords used is encrypted? On the scale of 1-5, with 5 being 'strongly agree', 23% employees choose '5', where as 24% choose scale '4'

J. Findings and Conclusion:

Data security largely depends on combinations of different security measures. Security policies and awareness plays a very key role among other measures. Through this research several data security measures were investigated. Finding suggest that implementing proper security measures is getting more crucial especial in a networked environment. Further findings and conclusions are discussed below.

1) A formal physical restriction to data was found to be present in more than 70% of the company's departments. This imply that the companies management was giving due importance to data security. However there were some lapse i.e. from the findings it shows that the company did not have the right procedure on how to handle secure distribution and disposal of back up.

2) Finding suggested that more than 90 % of the departments used authentication (user name and password). The users believed use of unique username and password was final to securing data. All computers on the network had password which was not encrypted. From the findings it shows that when an employee moves from one department to the other 70% of the accounts and password remained intact. 72% of those interviewed disclosed that user accounts were not reviewed on a regular basis to remove unknown ones. All accounts that were inactive were not automatically disabled in the system after a predefined period.

3) Virus protection as a data security measures: all computers in the company were loaded with antivirus software. 77 % of staff contacted, were a ware of some of the signs that could indicate computer had contracted a virus. The finding shows that more than 60% of the staff was not aware of what they could do if they suspect that there computer had a virus. More than 60 % of staff contacted reported that there storage devices, software or data that originated from outside were not scanned or checked for virus prior being used. From the findings it showed that the work of updating and ensuring that the computer was up to date was left for the IT security officer. Formalization of all information security policies, including policies for access control, application and system development, operational, network and physical security formally were not documented in more than 80% of the departments. Awareness and training program was not part of company Security policy.

K. Guidelines for Mumias Sugar Company Security Rule Compliance

The research presented above leads to a discussion of data security and security measures that are appropriate for the MSC. Data security measures refer to different data security mechanisms based upon the companies view of its data, values of the data and users of the data. Value of the data is determined by the company. Data to be secure requires various types of control measures. The common problem across the board is the ownership of the data and its privacy. The various control measures all relates to companies management. This part includes development of guidelines which form the framework to support this research. Various security measures that have been explored is mainly because of the value the companies data has. Administrator builds data protection; ICT technical staff maintains service levels and companies systems because of the company's data. Maximizing data security at MSC requires company involvement for which we present the following set of guidelines

L. Recommended practices

The following list identifies all the practices that should be implemented. Considerations for Technical Solutions, provides discussion regarding selected technical solutions.

- Account Management
- Information Management
- Disaster Recovery
- Electronic Mail
- Data Centers
- Remote Access
- Information for Users
- Workforce Identity
- Continuity Planning

M. Workforce Identity and Account Management

1) Determine which individuals are authorized to work with the network computer in the company in accordance with a role-based access approach.[164.308.a.3] [A]

2) Establish data security and control measures training for all members of the company workforce who are involved in the creation, transmission, and storage of the company's data. Ensure that training program includes periodic security reminders and is updated to take into account current vulnerabilities and threats. [164.308.a.5][a]

3) Take disciplinary action in accordance with MSC personnel policies and guidelines on workforce members who fail to comply with MSC policy and procedures, including information security policy and procedures.

4) Ensure the verification of the individual or employee who is authorized to access MSC system and that the person is correctly bound to a unique user identification ("sign-in") for access to the system [164.308.a.4][A] [164.312.a.1][R]

5) Ensure appropriate access controls mechanisms for authorized users' access to any MSC system. For systems with the very sensitive data, require strong electronic authentication, such as sufficiently complex passwords or use of other encryption key mechanisms to access the companies systems containing data. [164.308.a.5][A].

6) Establish account maintenance procedures that ensure termination of accounts or change in access privileges for individuals or entities who have terminated or no longer are authorized to access MSC systems [164.308.a.4][A].

7) Carefully manage system administrator accounts to ensure the accounts are used for only specific system administration functions. The number of these accounts should be kept to a minimum and provided only to personnel authorized to perform identified functions. Passwords or other authentication measures should be changed upon the termination of systems personnel who accessed these accounts.

8) Log activities performed by system administrator accounts and monitor logs on a regular basis.

REFERENCE

[1] Butler, S. A.(2002) "Security Attribute Evaluation Method: A Cost-Benefit Approach,"

[2] Geer, D., Soo Hoo, K., J., Jaquith, A.(2003) "Information Security: Why the Future Belongs to Quants," IEEE Security and Privacy.

[3] Mitnick, K., & Simon, W.(2002) The art of deception: Controlling the human element of security. Wiley Publishing.

[4] Siponen, M.(2001) Five dimensions of information security awareness. Computers and Society, June 2001, 24-29.

[5] Anderson J. (1972) "Computer Security Technology Planning Study," U.S. Air Force Electronic Systems Division Tech. Rep.

[6] Cohe F(1991)., "A Cost Analysis of Typical Computer Viruses and Defenses,"Computers & Security.

[7] Sandhu, R. S.(1996) Coyne, E., J., Youman, C. E., 1996, "Role-based Administration of Rules," ACM Transactions of Information Systems.

[8] Berinato, S. (2007). The end of innocence. CIO Magazine. Retrieved, from http://www.cio.com/article/133600/

[9] Johnson (2008). Information risk of inadvertent disclosure: Journal of Management Information Systems.

[10] Kumar & Park (2008). Understanding the value of countermeasures portfolios in information systems security

[11] Nyanchama, M. (2005). Enterprise vulnerability management and its role in information security management.

RC4 stream cipher and possible attacks on WEP

Lazar Stošić
College for professional studies educators
Aleksinac, Serbia

Milena Bogdanović
Teacher Training Faculty
University of Niš
Vranje, Serbia

***Abstract*—In this paper we analyze and present some weaknesses and possible attacks on the RC4 stream cipher which were published in many journals. We review some advantages and disadvantages which come from several authors, as well as similarities and differences which can be observed in the published results. Also, we analyze the Key Scheduling Algorithm (KSA) which derives the initial state from a variable size key, and strengths and weaknesses of the RCS stream cipher. Using examples from other papers, we show that RC4 is completely insecure in a common mode of operation which is used in the widely deployed Wired Equivalent Privacy protocol (WEP, which is part of the 802.11 standard).**

Keywords-RC4 stream cipher; KSA; WEP; security of WEP; WEP attack.

I. INTRODUCTION

RC4, a fast output-feedback cipher, is one of the most widely used cryptosystems on the Internet, commonly used as the default cipher for SSL/TLS connections [20]. It was designed by Ron Rivest in 1987 for RSA Data Security, Inc., and kept as a trade secret until it leaked out in 1994 and is now available for public analysis [18]. RC4 is currently being standardized by the IETF under the name "Arcfour" [23]. RSA DSI did not confirm that the published algorithm is in the RC4 algorithm, but experimental tests showed that it produces the same outputs as the RC4 software. The RC4 key stream generation algorithm updates the RC4 internal state and generates one byte of key stream. The key stream is XORed to the plaintext to generate the ciphertext. RC4 is comprised of two algorithms: the Key Scheduling Algorithm (KSA) which turns a random key (whose typical size is 40-256 bits) into an initial permutation S of {0,..., N-1}, which uses the secret key to create a pseudo-random initial state, and the Pseudo Random Generation Algorithm (PRGA), which generates the pseudo-random stream to generate a pseudo-random output sequence. Both algorithms are presented in Fig. 1.

KSA(K)	PRGA(S)
Initialization:	Initialization:
For $i = 0$ to $N - 1$	$i \leftarrow 0$
$S[i] = i$	$j \leftarrow 0$
$j \leftarrow 0$	Generation loop:
Scrambling:	$i \leftarrow i + 1$
For $i = 0$ to $N - 1$	$j \leftarrow j + S[i]$
$j \leftarrow j + S[i] + K[i \bmod l]$	Swap($S[i], S[j]$)
Swap($S[i], S[j]$)	Output $S[S[i] + S[j]]$

Figure 1. The RC4 Algorithms

(The Key Scheduling Algorithm and the Pseudo-Random Generation Algorithm)

In practical applications, stream ciphers are used with a session key which is derived from a shared secret key and an Initial Value (IV, which is transmitted unencrypted). The derivation of the session key can be done in various ways, such as concatenated ager the IV.

Section I is the introduction to this paper. Section II presents the features of RC4 family ciphers, strengths and weaknesses of the RC4 stream cipher and existing attack methods aimed at them.

Section III shows the Wired Equivalent Privacy protocol, used for encrypting wirelessly transmitted packets on IEEE 802.11 networks.

Section IV presents discussion of what this study has shown, strengths and weaknesses of the methods, how the results support the current literature or refute current knowledge and their impact on current thinking or practice. Section V concludes this paper.

II. RC4 STREAM CIPHER

RC4 has a secret internal state which is a permutation of all N=2n possible n bits words, along with two indices in it. In practical applications, n=8, and thus RC4, has a huge state of (1)

$$\log_2([[256]]^2 \cdot |S_{256}|) = \log_2(2^{16} \cdot 256!) \approx 1700 \ \beta\iota\tau\sigma \quad (1)$$

The initial state is derived from a variable-size key by a Key-Scheduling Algorithm (KSA), and then RC4 alternately modifies the state (by exchanging two out of the N values) and produces an output (by picking one of the N values).

RC4's internal state consists of a 256-byte array S, defining a permutation, as well as two integers $0 \leq i; j \leq 255$ acting as pointers into the array.

The RC4 key setup initializes the internal state using a key K of up to 256 bytes. RC4 keys are 2048 bits long, and their internal state consists of two counters i and j (each within □0□255_) plus an array of 256 8-bit bytes, called the S-box.

The S-box is initialized using the key K as follows (2):

```
for  i = 0 to 255
        S(i) = i
        j = 0

for  i = 0 to 255
        j = (j + S(i) + K(i)) mod 256
        swap S(i) and S(j)

i = 0
j = 0
```

(2)

Each next byte b of the keystream is produced using (3):

```
i = (i + 1) mod 256
j = (j + S(i)) mod 256
swap S(i) and S(j)
b = S(S(i) + S(j)) mod 256
```

(3)

For shorter key, the key is repeated as many times as necessary to fill the 2048-bit key. Once the S-box is initialized with the key, the RC4 algorithm is a loop that updates the internal state of the S-box and returns a byte of keystream. RC4 only protects the secrecy of a message, not its integrity. Other measures, such as the use of cryptographic checksums, are commonly used along with RC4.

RC4 stream cipher is used to protect internet traffic as part of the SSL (Secure Socket Layer) and TLS (Transport Layer Security) protocols, and to protect wireless networks as part of the WEP (Wired Equivalent Privacy) and WPA (Wi-Fi Protected Access) protocols. This attack was described by Fluhrer, Mantin and Shamir. It is a symmetric key algorithm and it is an important class of encryption algorithms. They encrypt individual characters (usually binary digits) of a plaintext message one at a time, using a simple time-dependent encryption transformation. The Blum-Goldwasser probabilistic public-key encryption scheme described in [7] is an example of asymmetric stream cipher. The same algorithm is used for both encryption and decryption as the data stream is simply XORed with the generated key sequence. The key stream is completely independent of the plaintext used. It uses a variable length key from 1 to 256 bits to initialize a 256-bit state table. The state table is used for Subsequent generation of pseudo-random bits and then to generate a pseudo-random stream which is XORed with the plaintext to give the cipher text.

The steps for RC4 encryption algorithm are as follows [17] and [16]:

1) Get the data to be encrypted and the selected key.
2) Create two string arrays.
3) Initiate one array with numbers from 0 to 255.
4) Fill the other array with the selected key.
5) Randomize the first array depending on the array of the key.
6) Randomize the first array within itself to generate the final key stream.
7) XOR the final key stream with the data to be encrypted to give cipher text.

One of the weaknesses of RC4 initialization mechanism is a major statistical bias in the distribution of the first output words. This bias makes it trivial to distinguish between several hundred short outputs of RC4 and random strings by analyzing their second word. This weakness can be used to mount a practical ciphertext-only attack on RC4 in some broadcast applications, in which the same plaintext is sent to multiple recipients under different keys. This unique statistical behavior is independent of the KSA, and remains applicable even when RC4 starts with a totally random permutation.

RC4 Strengths:

Some of RC4 Strengths [16]:

1) The difficulty of knowing which location in the table is used to select each value in the sequence.
2) A particular RC4 key can be used only once.
3) Encryption is about 10 times faster than DES.

RC4 Weaknesses:

Some of RC4 weaknesses [17] and [16]:

1) The RC4 algorithm is vulnerable to analytic attacks of the state table [6] and [15].
2) WEAK KEYS: these are keys identified by cryptanalysis that is able to find circumstances under which one or more generated bytes are strongly correlated with small subset of the key bytes. These keys can happen in one out of 256 keys generated [10], [5] and [15].

Many ways to break RC4 are classified as Distinguishing Attack, which makes use of the bias in output sequence. In 2004, some new stream ciphers were proposed, to which resistance to the attacks aimed at RC4 was added. They are exemplified by VMPC, a stream cipher proposed by B. Zoltak [4], and RC4A, an RC4 family algorithm improved by S. Paul and B. Preneel [11].

III. WIRED EQUIVALENT PRIVACY

Today, PC cards are most frequently used in home and business networks. All computers have a security protocol called Wired Equivalent Privacy (WEP). A device using an 802.11 card is configured with a key, that in practice usually consists of a password or a key derived from a password.

Wired Equivalent Privacy (WEP) is a protocol for encrypting wirelessly transmitted packets on IEEE 802.11 networks. In a WEP protected network, all packets are encrypted using the stream cipher RC4 under a common key, the root key1 Rk. Rk is the WEP or root key and IV is the initialization vector for a packet. $K = Rk \| IV$ is the session or per packet key. X is a key stream generated using K. The WEP protocol is designed to provide privacy to packet based wireless networks based on the 802.11b standard [19]. The WEP encrypts by taking a secret key and a per-packet 3 byte IV, and using the IV followed by the secret key as the RC4 key. The attacker is able to retrieve the first byte of the RC4 output from each packet.

WEP uses a 40-bit secret key (which was the largest easily exportable key when WEP was designed), shared between all the users and the network access point. For every packet, the sender chooses a new 24 bit Initialization Vector (IV), and the 64-bit RC4 key is the concatenation of the chosen IV (occurring first) and the shared key (occurring last). Such an IV-based mode of operation is commonly used in stream ciphers in order to generate different PRGA outputs from the same long term key, and the frequent resetting of the PRGA is designed to overcome the unreliable nature of the Wireless LAN environment.

The simplest weakness is the small size of the secret key and the IV: A 40-bit key can be recovered by an exhaustive search in less than one day. The limited size (224) of the IV space implies that IVs are reused during the encryption of different packets. This mode can be attacked by constructing a dictionary of all the 224 IVs along with their corresponding key streams. WEP defines no easy mechanism for changing the shared key, and thus the key is usually changed only infrequently, increasing the attacker's chance to construct this dictionary.

The first "real" attack makes it possible to derive an arbitrarily long key in time which grows only linearly with its length in the weakest attack model of known plaintext and IV developed in [12], and is outlined in the next section.

A first analysis of the design failures of the WEP protocol was published by Borisov, Goldberg and Wagner [9] in 2001, which showed that the IV merely protects against random errors but not against malicious attackers. They observed that old IV values could be reused, thus allowing to inject messages.

RC4 are specified. Several PC cards reset IVs to zero every time they are initialized, and then increment them by one for every use. This results in high likelihood that keystreams will be reused, leading to simple cryptanalytic attacks against the cipher, and decryption of message traffic.

Fluhrer, Mantin and Shamir presented a related key ciphertext-only attack against RC4 [13] as used in WEP. In WEP, the key scheduling algorithm uses either a 64-bit packet key (40-bit secret key plus 24-bit IV) or a 128-bit key (104-bit secret key plus 24-bit IV) to set up the RC4 state array, S, which is a permutation of {0,…, 255}. The output generator uses the state array S as well as two counters, i and j, to create a pseudorandom sequence.

In order for this attack to work, the IVs need to fulfill a so-called "resolved condition". This attack was suspected to be applicable to WEP, which was later demonstrated by Stubblefield et al. [1]. Approximately 4 million different frames need to be captured to mount this attack. Vendors reacted to this attack by filtering IVs fulfilling the resolved condition, so-called "weak IVs". On the other side, Klein [3] showed an improved way of attacking RC4 using related keys that does not need the "resolved condition" on the IVs and gets by with a significantly reduced number of frames.

A. The WEP Attack

The attack starts with the known IV as a basis, and repeatedly applies the sub-attack in order to recover all the keywords in the secret key SK. To conduct an attack, the cryptanalyst needs the first output word of a large number of RC4 streams along with the IV that was used to generate each one of them. Since in WEP the IVs are transmitted in the clear and the first message word in most packets is a known constant, these requirements are automatically satisfied.

With about 60 such IVs, the attacker can re-derive the key byte with reasonable probability of success. The number of packets required to obtain that number of IVs depends on the exact IVs that the sender uses. Although the 802.11b standard does not specify how an implementation should generate these IVs, common practice is to use a counter to generate them. We now analyze the performance of this attack for two different counter modes. If the counter does not start from zero, the attacker has an alternative strategy available to him. If the attacker assumes the first two bytes of secret key, than for each initial IV byte, there are approximately 4 settings of the remaining two bytes that set up the permutation as required to re-derive a particular key byte.

Fluhrer S., Mantin I. and Shamir A. in their work Attacks on RC4 and WEP explain that the first x words of the KSA key are known. This makes it possible to simulate the first x rounds of the KSA and compute the permutation S_{x-1} and the indices i_{x-1} and jx-1 at that point. The next value of i is also known (i_x=x) but the next value of $j(j_x)$ depends on the unknown target keyword K[x] (since $j_x=j_x-1+S_x-1[x]+K[x]$) and thus each of the values j_x and K[x] can be easily derived from the other. Consequently, given $S_x[x]$, we can compute which value was in position j_x in the known permutation S_{x-1}, and by inverting this permutation, we can recover j_x itself.

IV. DISCUSSION

RC4 is a symmetric key algorithm. Stream cipher is an important class of encryption algorithms. They encrypt individual characters of a plaintext message one at a time, using a simple time-dependent encryption transformation. RC4 is comprised of two algorithms: the Key Scheduling Algorithm (KSA) which turns a random key (whose typical size is 40-256 bits) into an initial permutation S of {0,..., N-1}, which uses the secret key to create a pseudo-random initial state, and the Pseudo Random Generation Algorithm (PRGA), which generates the pseudo-random stream to generate a pseudo-random output sequence.

We see that some of RC4 strengths were: the difficulty of knowing which location in the table is used to select each value in the sequence; a particular RC4 key can be used only once; encryption is about 10 times faster than DES. On the other side, RC4 weaknesses were: The RC4 algorithm is vulnerable to analytic attacks of the state table; WEAK KEYS: these are keys identified by cryptanalysis that is able to find circumstances under which one or more generated bytes are strongly correlated with small subset of the key bytes.

In current literature we can see that many ways to break RC4 are classified as Distinguishing Attack. These ways make use of the bias in output sequence. The first "real" attack makes it possible to derive an arbitrarily long key in time which grows only linearly with its length in the weakest attack model of known plaintext and IV. A first analysis of the design failures of the WEP protocol showed that the IV merely protects against random errors but not against malicious attackers. They observed that old IV values could be reused, thus allowing to inject messages.

In the last few decades many stream ciphers have been proposed. Most of them are easy to implement on hardware but their performance is slow when implemented on software. Since RC4 is such a widely used stream cipher, it attracted considerable attention in the research community since it was proposed. The strength of the RC4 key does not grow linearly with the increase in the key length.

V. CONCLUSIONS

The main contribution of this paper is the presentation of the established and proven deficiencies of RC4 which are caused by its extreme simplicity. Based on the results of numerous research studies, we can conclude that the initialization of the pseudo-random index j to 0 seems to be the most problematic operation, and both the second byte bias and the IV weakness could be avoided by using a more complex initialization of j. Possible methods for initializing j are to use j from the end of the KSA or to give it the value of one of the key words. The invariance weakness and the IV weakness are inherent consequences of the structure of the KSA.

For the RC4 stream cipher, every key has a family of related keys which result in a substantially similar keystream. The strength of the RC4 key does not grow linearly with the increase in the key length. If RC4 is deployed using keys longer than the customary 128 bits, we advise discarding the first 256 bytes of the keystream.

A perfect initialization mechanism is not easy to achieve. A common mode of operation to achieve these contradicting goals is to discard a prefix of output bits. These mute rounds usually disconnect the generated stream from the initialization process, and improve the "randomness" of the generated stream.

The discarded prefix should also grow in the same way (exponentially) when enlarging RC4 words into 16 bits (which is sometimes recommended for faster encryption of large amount of data). The expression of the invariance weakness spreads over several hundred words in RC416 and eliminating only 256 words is not sufficient when N is larger. The reduced version RC46 can be attacked with practical complexity, while for stronger version (RC4n>6) it is possible to mount enhanced (but impractical) attacks.

REFERENCES

[1] A. Stubble, J. Ioannidis, and A. D. Rubin, "A key recovery attack on the 802.11b wired equivalent privacy protocol (WEP)", ACM Transactions on Information and System Security, Volume 7 Issue 2, May 2004, pp. 319-332.

[2] A. Bittau, M.Handley, and J. Lackey, "The final nail in WEP's coffin. In IEEE Symposium on Security and Privacy", pp. 386-400, IEEE Computer Society, 2006.

[3] A. Klein, "Attacks on the RC4 stream cipher", volume 48 Issue 3, pp. 269 – 286, Designs, Codes and Cryptography, September 2008. doi>10.1007/s10623-008-9206-6

[4] B. Zoltak: "VMPC One-Way Function and Stream Cipher," Fast Software Encryption, FSE 2004, LNCS 3017, pp.210-225, Springer-Verlag, 2004.

[5] E. Biham and Y. Carmeli, "Efficient Reconstruction of RC4 Keys from Internal States", FSE 2008, pp. 270-288, vol. 5086, Lecture Notes in Computer Science, Springer.

[6] G. Paul, S. Rathi and S.Maitra, "Non-negligible Bias of the First Output Byte of RC4 towards the First Three Bytes of the Secret Key", Proceedings of the International Workshop on Coding and Cryptography (WCC) 2007, pp. 285-294 and Designs, Codes and Cryptography Journal, pp. 123-134, vol. 49, no. 1-3, December 2008.

[7] M. Biriyukov, A. Shamir, and D. Wagner, "Real time cryptanalysis of A5/1 on a PC", FSE: Fast Software Encryption, 2000., pp. 1-18.

[8] M. Akgun, P.r Kavak, H. Demirci, "New Results on the Key Scheduling Algorithm of RC4", INDOCRYPT 2008, pp. 40-52, vol. 5365, Lecture Notes in Computer Science, Springer.

[9] N. Borisov, I. Goldberg, and D. Wagner, "Intercepting mobile communications: the insecurity of 802.11", In ACM MobiCom 2001, pp. 180-189. ACM Press, 2001.

[10] R. Basu, S. Maitra, G. Paul and T. Talukdar, "Some Sequences of the Secret Pseudo-random Index j in RC4 Key Scheduling", Proceedings of the 18th International Symposium on Applied Algebra, Algebraic Algorithms and Error Correcting Codes (AAECC), June 8-12, 2009, Tarragona, Spain, pp. 137-148, vol. 5527, Lecture Notes in Computer Science, Springer.

[11] S. Paul, and B. Preneel, "A New Weakness in the RC4 Keystream Generator," Fast Software Encryption, FSE 2004, LNCS 3017, pp.245-259, Springer-Verlag, 2004.

[12] S. R. Fluhrer, I. Mantin, and A. Shamir, "Weaknesses in the key scheduling algorithm of RC4", In SAC`2001, 2001.

[13] S. R. Fluhrer, I. Mantin, and A. Shamir, "Weaknesses in the key scheduling algorithm of RC4", In Serge Vaudenay and Amr M. Youssef, editors, Selected Areas in Cryptography 2001, volume 2259 of Lecture Notes in Computer Science, pp. 1-24. Springer, 2001.

[14] S. Dorhofer, "Empirische Untersuchungen zur WLAN-Sicherheit mittels Wardriving", Diplomarbeit, RWTH Aachen, September 2006. (in German).

[15] V. Tomašević, S. Bojanić, O. Nieto-Taladriz, "Finding an internal state of RC4 stream cipher", Information Sciences, Volume 177, issue 7, 01. April, 2007, pp.1715-1727.

[16] W. Mao, Modern Cryptography Theory and Practice, Prentice Hall, New Jersey, 2004.

[17] W. Stilings, Cryptography and Network Security Principles and practices, Fourth Edition, PEARSON, USA, 2006.

[18] B. Schneier, Applied Cryptography, John Wiley and Sons, New York, 2nd edition, 1996.

[19] LAN/MAN Standard Committee, Wireless LAN medium access control (MAC) and physical layer (PHY) specifications, 1999 edition, IEEE standard 802.11, IEEE Computer Society, 1999.

[20] T. Dierks and C. Allen, The TLS Protocol, Version 1.0, Internet Engineering Task Force, January 1999.

[21] E. Tews, R. P. Weinmann and A. Pyshkin, "Breaking 104 bit WEP in less than 60 seconds", IACR Eprint Server, http://eprint.iacr.org/2007/120.pdf , number 2007/120, Accessed April 1, 2007.

[22] S. Fluhrer, I. Mantin, A. Shamir, "Attacks on RC4 and WEP, RSA Laboratories", http://www.rsa.com/rsalabs/cryptobytes/cryptobytes_v5n2.pdf Volume 5, No. 2, Summer/Fall 2002

[23] K. Kaukonen and R. Thayer, "A Stream Cipher Encryption Algorithm Arcfour", http://tools.ietf.org/html/draft-kaukonen-cipher-arcfour-03 , Internet Engineering Task Force (IETF), July 1999.

AUTHORS PROFILE

Lazar Stošić received the Ph.D. degrees of computer sciences on Faculty of Informatics and Information Technology, Novi Pazar, Serbia. Professor on College for professional studies educators, Aleksinac, Serbia. Member of the Society for media and science, e-learning center at the University of Zurich since 2009. - Gesellschaft für Medien in der Wissenschaft (GMW) e. V., E-Learning Center der Universität Zürich, Zürich. Member of the Society for Computer Science, Germany since 01.07.2009. - Gesellschaft für Informatike e.V. (GI), German Informatics Society, Mitgliedsnummer: GI 59631 AhrstraBe 45, 53175 Bonn, Germany. His research interests include Information Technology, Computer System, Computer Education and Computer Distance Learning.

Milena Bogdanović is the assistant professor at the Theacher Training Faculty in Vranje, Serbia (major in Mathematics and Informatics – Mathematics 1, Mathematics 2, Elementary mathematical concepts, IT in Education, Educational technology, Elements of mathematics). She is the Reviewer of international journals - IJACSA – International Journal of Advanced Computer Science and Applications; Member of the editorial boards of international journals – International Journal of Computer Systems and Applications – Inernational Scientific Press; Reviewer of the Book of solved tasks in Mathematics 2; Experience in teaching in secondary school and university; Participant in numerous seminars and training for eduational reform, active learning, Lifelong Learning, Mathematics and Applications. She is the author of two books and of about 40 of published scientific papers in the field of the mathematics and computer science. Her professional papers discuss problems in the field of applications of multimedia in teaching, combinatorial optimization, genetic algorithms, directable automata.

Transforming Conceptual Model into Logical Model for Temporal Data Warehouse Security: A Case Study

Marwa S.Farhan
Information Systems Dep,
Faculty of Computers and Information, Helwan University,
Cairo, Egypt

Mohamed E. Marie
Faculty of Computers and Information, Helwan University,
Cairo, Egypt

Laila M. El-Fangary & Yehia K. Helmy
Faculty of Computers and Information, Helwan University,
Cairo, Egypt

***Abstract*—Extraction–transformation–loading (ETL) processes are responsible for the extraction of data from several sources, their cleansing, customization and insertion into a data warehouse. Data warehouse often store historical information which is extracted from multiple, heterogeneous, autonomous and distributed data sources, thereby, the survival of the organizations depends on the correct management, security and confidentiality of the information. In this paper, we are using the Model Driven Architecture (MDA) approach to represent logical model requirements for secure Temporal Data Warehouses (TDW). We use the Platform-Independent Model (PIM) which does not include information about specific platforms and technologies. Nowadays, the most crucial issue in MDA is the transformation between a PIM and Platform Specific Models PSM. Thus, OMG defines use the Query/View/Transformation (QVT) language, an approach for expressing these MDA transformations. This paper proposes a set of rules to transform PIM model for secure temporal data warehouse (TDW) to PSM model, we apply the QVT language to the development of a secure data warehouse by means of a case study.**

Keywords- ETL; temporal data warehouse; Data warehouse security; MDA; QVT; PIM; PSM.

I. INTRODUCTION

Data warehouse often store historical information which is extracted from multiple, heterogeneous, autonomous and distributed data sources, thereby; the survival of the organizations depends on the correct management, security and confidentiality of the information. The application of the Model Driven Architecture (MDA) [1] in the secure modeling of DWs allows obtaining the secure logical scheme from the conceptual model. In this work we apply a set of QVT [2] relations to the development of a secure DW. Various approaches for the conceptual design of the DW repository have been proposed in [4, 5, 6, 7]. These proposals are twofold; on the one hand they try to represent the main MD properties at the conceptual level by abstracting away details of the target database platform where the DW will be implemented. On the other hand, they also define how to derive a logical representation tailored to a specific database technology (relational or multidimensional). These approaches are lacking in formal mechanisms to univocally and automatically obtain the logical representation of the conceptual model [3].

In order to overcome this limitation, in previous proposals, we have described a model driven framework for the development of DWs, based on the MDA standards. In this paper we will propose a set of rules to transform the proposed conceptual model in [8] to logical model, our model include two stages (1): ETL stage we use UML class diagram to represent ETL processes in the logical model. (2):DW stage we use the Query/View/Transformation (QVT) language [2] to the MD modeling of the DW repository within our MDA framework. The PIM model can be translated into: (1) one or more Platform Specific Models (PSM) with information about the specific technology used; or (2) other PIMs with a different level of abstraction. Each PSM can then be translated into a code that can be executed in the specific platform. The proposed model focuses on the logical modeling for temporal Data warehouse (TDW) considering the DW security issues. This paper is organized as the following: section 2 presents the related work. Section 3 presents ETL logical model considerations, Section 4 describes MDA and QVT features and an example is provided in this section to show how to apply MDA and QVT transformation rules. Finally, section 5 points out our conclusions and future works.

II. RELATED WORK

This section divides the related work according to two main research topics covered by this paper: *ETL modeling*, and *data warehouse modeling*. The paper focuses on the logical modeling specifications in these topics.

A. ETL Modeling

The modeling and optimization of ETL processes at the logical level is presented in [9], [10]. The authors of [11] proposed a design method that includes an algorithmic transformation of conceptual to logical models for ETL processes. The conceptual modeling of the ETL processes is discussed in [12]. In [13, 14] the authors focus on the dynamic [13] and static [14] modeling of the ETL processes. There are few research papers on DW security and OLAP (Online Analytical Processing) security [15, 16, 17, 18]. But none of the above mentioned papers discuss an integrated security model for ETL processes. In this paper, we propose the security aspects and temporal aspects that have to be considered in the

analysis and design phases of ETL processes and DW. Supporting our proposal the lack of security issues in ETL are mentioned in [19, 20].

B. Data warehouse Modeling

There are interesting contributions in the field of information systems security but they do not deal with DWs in the context of their specific security issues. One of the most relevant proposals that integrate security through the use of UML is UMLsec [21], which can be used to specify and evaluate UML security specifications using formal semantics. Furthermore, Model-Driven Security (MDS) [22] extends MDA to build secure information systems. Its designers specify the inclusion of security properties in high-level system models and use tools to automatically generate secure system architectures. Within the context of MDS, the same authors propose an extension of UML for modeling a generalized Role-Based Access Control (RBAC) called SecureUML [23].

Data Warehouses (DWs) present specific characteristics and security challenges related to all their layers and operations [24]. Proposals for DWs at conceptual and logical levels, which consider special characteristics of DWs, also exist, but they do not support security issues. The most interesting proposal is [25] in which the authors define a methodology to analyze security requirements, to represent them at the conceptual level .However; they do not define the transformation between levels. Another important proposals are [20,26,27,30,32]which provides security models at different abstraction levels and has been aligned with an MDA architecture in which security models are embedded and scattered throughout the high-level system models, which are transformed towards the final implementation according to the MDA strategy. However these models are consider the read operation in DW and didn't consider the temporal DW requirements. Our research efforts are thus applied to the development of secure DWs considering confidentiality issues during the whole development process, from an early development stage to the logical model. Our scope in this paper is how to transform conceptual model to logical model considering DW security and temporal issues.

III. ETL Logical Model Considerations

During the ETL process, data is extracted from an OLTP databases, transformed to match the data warehouse schema, and loaded into the data warehouse database. Fig. 1 shows the general framework for ETL processes. In the bottom layer we depict the data stores that are involved in the overall process. On the left side, we can observe the original data providers (typically, relational databases and files). The data from these sources are extracted (as shown in the upper left part of Fig. 1) by extraction routines, which provide either complete snapshots or differentials of the data sources. Then, these data are propagated to the *Data Staging Area* (DSA) where they are transformed and cleaned before being loaded to the data warehouse. The data warehouse is depicted in the right part of Fig. 1 and comprises the target data stores, i.e., fact tables and dimension tables. Eventually, the loading of the central warehouse is performed through the loading activities depicted on the upper right part of the figure [12].

The ETL process is not a one-time event. As data sources change the data warehouse will periodically updated. Also, as business changes the DW system needs to change in order to maintain its value as a tool for decision makers, as a result of that the ETL also changes and evolves. The ETL processes must be designed for ease of modification. A solid, well-designed, and documented ETL system is necessary for the success of a data warehouse project. As fig.1 shows An ETL system consists of three consecutive functional steps: extraction, transformation, and loading, in the following sections we will explain the ETL stage of our PSM. Our model transforms the PIM model which we proposed in [8] to PSM.

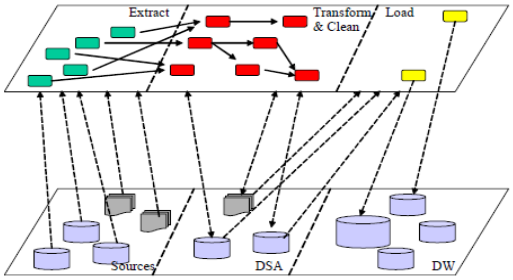


Figure 1.The environment of ETL processes

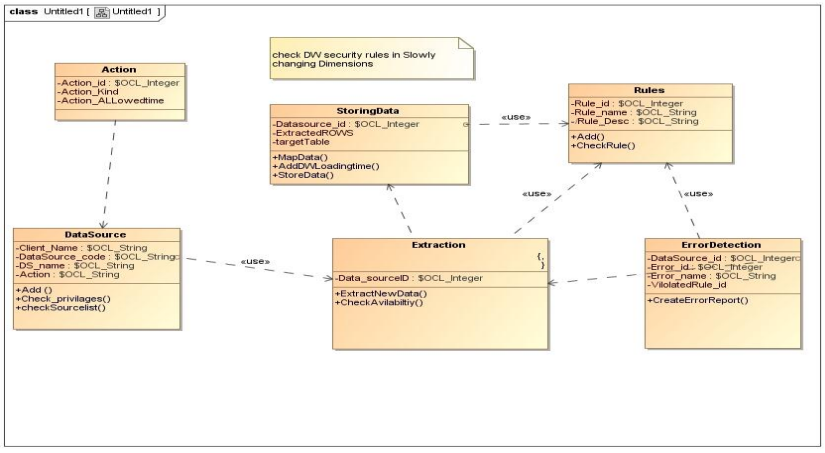


Figure 2. ETL logical model (ETL PSM)

We will use the UML class diagram because this is the most standardize way to express the ETL activities in the logicallevel.Fig.2 shows the ETL PSM model for the whole ETL processes as follows

A. Extraction

The first step in any ETL scenario is data extraction; in this step we extract data from their sources. We have 3 classes in this step:

1) DataSource class and Action Class (in parallel)

These two classes are working in parallel; the scenario is graphically depicted in Fig.2 and involves the following transformations.

1-first, we check the data source in the data source metadata and in the same time we check the allowed privileges in the action class for each data source.

2. Second, if we find any invalid privileges to any data source we reject the data extraction from this data source. We do this by comparing the required action in DS class and allowed action in Action class.

2) Extraction Class

3. After checking data source availability we use Extraction class. This class is responsible for extracting the modified and new data from data sources not extract all data. We do this by comparing the incoming data with the data in old Trans table which we stored in it the last refresh cycle data to extract the new data from the last DW refresh cycle, this table is stored in ETL metadata.

B. Transformation

The transformation step tends to make some cleaning and conforming on the incoming data to gain accurate data which is correct, complete, consistent, and unambiguous. This process includes data cleaning, transformation, and integration. *This stage includes three classes (Extraction class, Rule Class, Error detection), these classes are working in parallel*

Extraction class: as we explained earlier we extracted the modified data, in the transformation step we check these data against rules in Rule class.
Rule Class: we check data against set of privileges, these privileges contains (1)Security privileges which considers security rules which defined in (my paper);(2)Access privileges which includes:

a) ***Read access:*** *is the normal read operation it means just query data is allowed.*

b) ***Write access****: includes (insert) for temporal and non temporal data. insert, update or delete in case of temporal(Slowly Changing Dimension(SCD),these security constraints is explained in details in[8]*

Error detection class: this class is responsible for generate report with the invalid data using attributes and operations in Rule class to describe data source, the violated rules, error description, error time …etc.

C. Loading

Loading data to the target multidimensional structure is the final ETL step. In this step, extracted and transformed data is written into the dimensional structures actually accessed by the end users and application systems. In the load process we check the loaded data against DW security constraints, if there's any invalid data it can be added to error report as we explained earlier. If data is valid we add the data warehouse loading time(DWLT) to the SCD to represent the time when the new data is loaded in DW; this DWLT represents the (Start_time and end_time) of each row in SCD.

IV. MDA AND QVT FEATURES

QVT is an essential part of the MDA standard as a means of defining formal and automatic transformations between models. The QVT is a standard approach for defining formal relations between MOF compliant models.QVT consists of two parts: declarative and imperative. (1)Declarative part: provides mechanisms to define transformation as a set of relations that must hold between the model elements of a set of candidate models (source and target models).(2)Imperative part: defines operational mappings that extend the declarative part with imperative implementations when it is difficult to provide a purely declarative specification of a relation.

The proposed model focuses on the declarative part of QVT because the scope of the paper is the logical model requirements. This paper focuses on the relational layer of QVT which supports the specification of relationships that must hold between MOF models by means of a relations language. A relation is defined by the following elements:

-Two or more domains: each domain is a set of elements of a source or a target model. The kind of relation between domains must be specified: checkonly (C), i.e., it is only checked if the relation holds or not; and enforced(E), i.e., the target model can be modified to satisfy the relation.

– When clause: it specifies the conditions under which the relation needs to hold (i.e. precondition).

– Where clause: it specifies the condition that must be satisfied by all model elements participating in the relation (i.e. postcondition).

Defining relations by using the QVT language has the following advantages : (i) it is a standard language, (ii) relations are formally established and automatically performed, and (iii) relations can be easily integrated in an MDA approach[28].

A. Transformation from PIM model to PSM model

This sections explains the proposed rules to transform from PIM model to PSM, we will base on the approaches which was presented in [29, 30].Fig.3 illustrates the Secure Multidimensional MDA architecture [1]. On the left hand side the Secure Multidimensional conceptual scheme, i.e., SMD PIM is presented. By means of the transformation T1 we obtain the relational logical scheme, i.e., SMD PSM, represented in the centre of Figure 3. If we choose a SGBD that implements security aspects, then SMD PSM is transformed according to T2 into code for the target platform. This code is called the Secure Multidimensional Code (SMD Code). The Figure illustrates how the security constraint defined by means of a security Rule (represented as an UML note) is transformed

from the conceptual level to the logical level by employing T1, and later transformed into code with the T2 transformation.

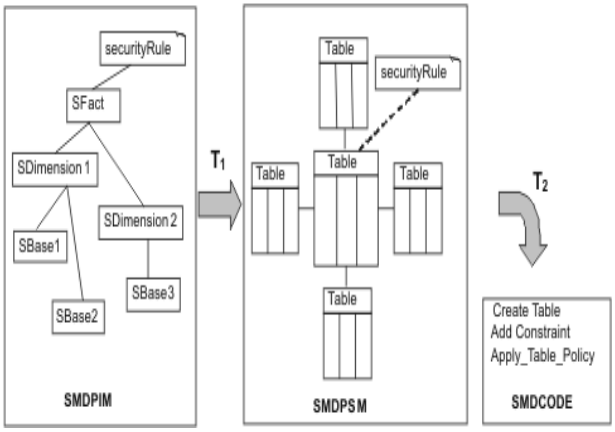


Figure 3.General Transformation Schema

B. QVT relations to obtain the PSM

In multidimensional modeling, the logical level is designed according to the specific properties of the SGBD (Relational Online Analytical Processing, ROLAP, Multidimensional Online Analytical Processing, MOLAP or Hybrid Online Analytical Processing, HOLAP). Still, Kimball [31] assures that the most common representation is on relational platforms, i.e., on ROLAP systems. The SMD PSM allows us to represent at the relational level the security requirements that were represented in the conceptual modeling of the DW. In this model we can represent tables, columns, primary and foreign keys, etc. Thus, we can establish security in attributes and tables. We express the security constraints that were modeled at the conceptual level by means of UML notes. We apply QVT declarative approach based on the proposed models in[27,29**]** **.**Fig.4 shows the case study which used in[29],we will use the same case study and we will use the QVT to represent the security requirements of TDW which was not considered in the QVT before this model.

Fig. 4 shows a secure MD model that includes a fact class (Admission), three dimensions classes (Diagnosis, Patient and Time), five base classes (DataD, Diagnosis_Group, DataP, City, and DataT), and a UserProfile class. The Admission fact class –SFact stereotype- contains all the individual admissions of patients in one or more hospitals, and can be acceded by all the users who have security levels secret or topSecret -labeled value SecurityLevels (SL)-, and perform health or administrative roles -tagged value SecurityRoles (SR)-. Be observed how the attribute cost only can be acceded by users whom play administratively role - tagged value SR- The class base DataP contains the information of the patients of the hospital and can be acceded by all the users who have security level secret - tagged value SL-, and play health or administrative roles -t value SR-.

The Address attribute can be only accessed by users who have an administrative role -tagged value SR of attributes-. City base class contains the information of cities, and it allows us to group patients by cities. City base class can be accessed by all users who have confidential security level -tagged value SL-. DataD base contains the information of each user diagnosis, and can be accessed by users who have a health role - tagged value SR-, and have secret security level -tagged value SL-. Finally Diagnosis_group contains a set of general groups of diagnosis. Each group can be related to several diagnoses, but a diagnosis will be always related to a group. Diagnosis_group can be accessed by all users who have confidential security level -tagged value SL-. Some security constraints have been specified by using the previously defined constraints.

1) Transforming SFacts into STables

The first relation in executing is SecureDW2SSchema, with it, all levels of security: confidential, secret and topsecret, as well as all the hierarchical roles tree is transformed into their equivalent ones of SSchema.

The UserProfile2RUserProfile relation transforms the UserProfile class into a table belonging to SSchema that will have the same name of UserProfile. The relation that follows, i.e., SFact2STable is shown in its graphical notation in Fig. 5, by means of this relation each SFact jointly with its security properties is transformed into a table that will contain the same information of security.

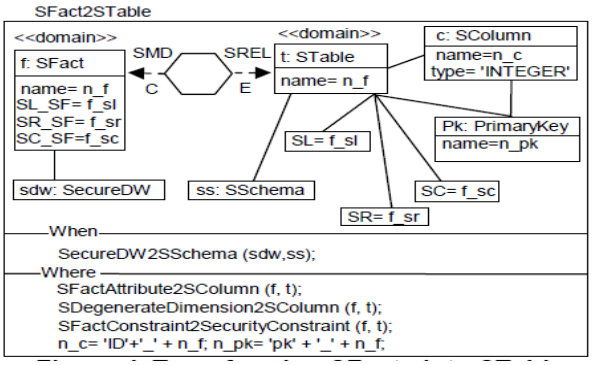


Figure 5 Transforming SFacts into STables

In Fig.6, we are going to show how the attributes of SFact are transformed into SColumns of the table that represents the SFact, so that, each column will contain the security information of its corresponding attribute in the SFact.

In Figure 7 we show the result of applying the SFact2STable relation to our case study. The SFact Admission is transformed into a table of the model SMD PSM, i.e., in the Admission table, that will have a primary key, as well as the security properties securityLevel and securityRole.

Fig. 8 shows the result of applying the SFactAttribute2SColumn relation, as a consequence, the Admission table will contain the columns respectively type and cost of type string and float. The column cost will have associated the security property securityRole. Also in Fig. 8, the associated requirements of security to the Admission table are modeled in the heading of the table, according to the SECRDW metamodel.

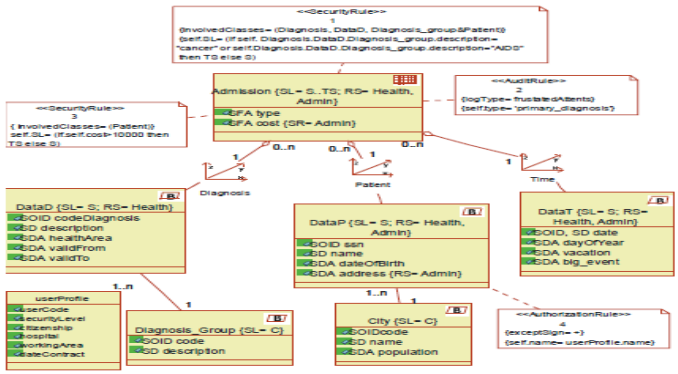


Figure 4: Example of secure multidimensional modeling

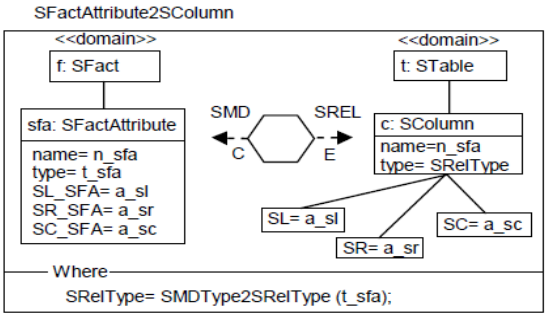


Figure 6 Transforming SFact attributes into Scolumn

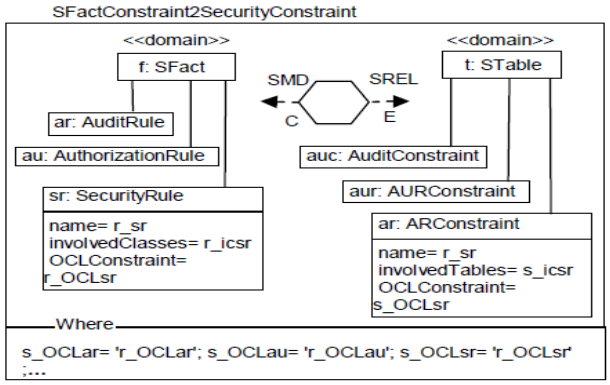


Figure 9 SFactConstraint2SecurityConstraint

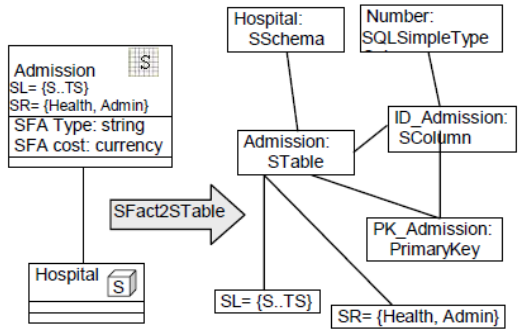


Figure 7 Applying SFact2Stable

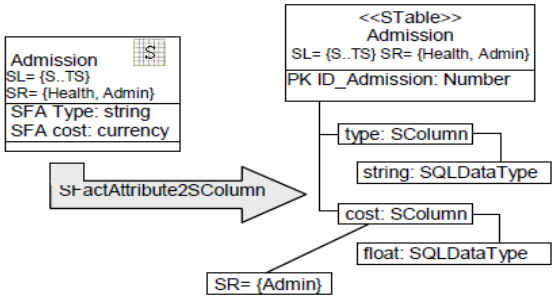


Figure 8 SFactConstraint2SecurityConstraint

Fig. 9 presents the definition of the SFactConstraint2SecurityConstraint relation, which guarantees that all the constraints associated with the SFact are transformed in constraints associated with the table, just as it can be seen in Fig. 10.

2) Transforming SDimension into Stable

In this section we will explain how to represent temporal and non-temporal dimensions.

Fig.10 we show the definition of the SDimension2STable relation. In multidimensional modeling the dimension do not have attributes [10]. For this reason, when the SDimension2STable relation is executed, a table is created whose name is merged with the names from dimension and rootBase respectively.

The rootBase is the only SBase associated with the SDimension. All the associated security information with the rootBase is transformed in security properties of the table and by means of the execution of the relations that appear in the clause *where* of the SDimension2STable relation, is guaranteed that all the attributes of the rootBase are going to conform the columns of the table.

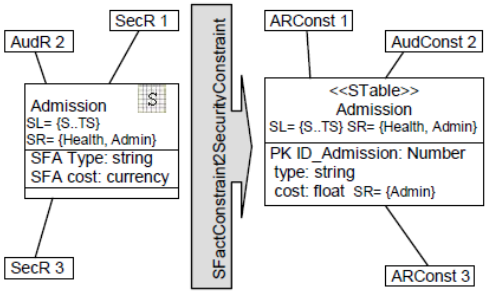


Figure 10 Applying SFactConstraint2SecurityConstraint

In the Fig. 11 we show the definition of the SBase2STable relation. This relation creates a table with a primary key, as well as a foreign key in the table that receives as parameter when it is invoked; logically the primary key and the foreign key will be associated for guaranteeing that the tables form a part of a relation one-to-many between the SBases. In the clause *where* this relation is called again, as well as the SpecializedSBase2STable relation to assure us that we cover the whole hierarchy of bases that conforms the dimension.

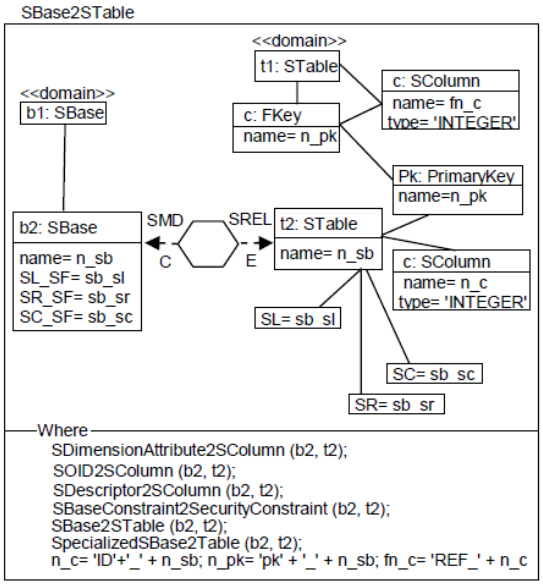


Figure 11 Transforming SBase into STable

In Fig.12 we illustrate the application of the SBase2STable relation to our case study. When the relation is executed, the City table is created with the primary key PK_City. This primary key will be associated with the foreign key that is also created in the STable Patient_Data. The City table will have associate the security property defined by means of securityLevel with confidential value. The final result is that the tables City and Patient_DataP are related.

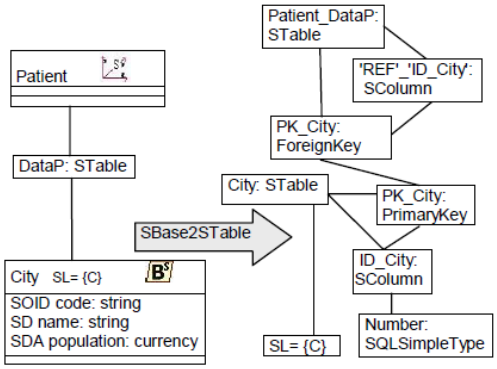


Figure 12 Applying SBase2STable

a) Dimension

In Fig. 13, fig. 14 we illustrate the applying of the SDimension2STable relation, as a result of applying this relation, the Patient_DataP table is created, with all the security properties that has associated the rootBase, in this case the security level secret and the user roles health and admin.Several of the relations that appear in the clause *where* from the SDimension2STable relation keep certain similarity with the defined ones for the SFact2STable relation, for that reason; next we are going to define the SBase2STable relation.

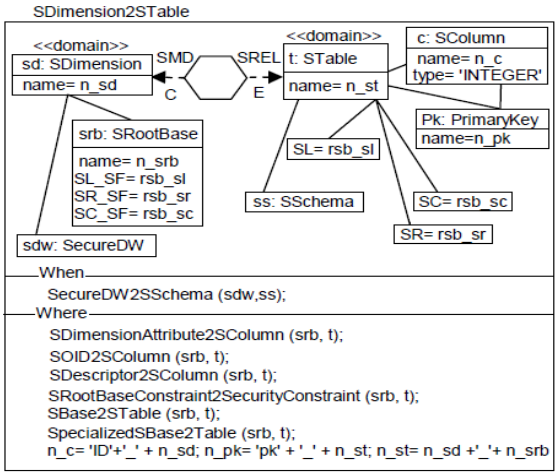


Figure 13 Transforming SDimension into STable

a) Slowly Changing Dimension2Stable

In Fig.15 we assumes that we have SCD in the schema, we illustrate how to apply QVT rules in case of we have SCD, we here assume that patient dimension is SCD,as a result of applying this relation, the Patient_DataP table is created, with all the security properties that has associated the rootBase, in this case the security level secret and the user roles health and admin.

We focus on temporal attributes as Valid time attribute (Stat_time,End_time) and dimension security privileges(DSP) that we explained earlier.

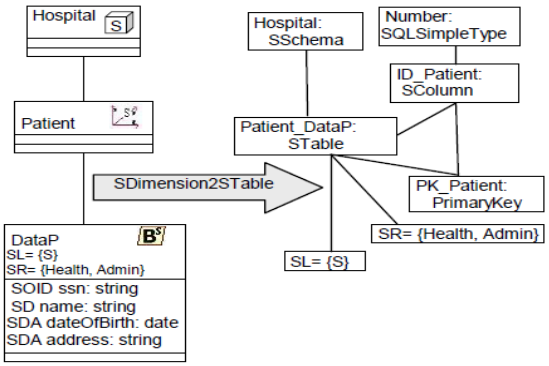


Figure 14 Applying SDimension2Stable

The dimension access security privileges (Insert,Upate,Delete) must be equal to BaseRoot access Privileges (Bsp).

In sake of the space we just clarify the SCD features and all the other features are the same in fig.13,fig.14.The first four relations that appears in the clause where of the SBase2STable relation guarantee the transformation of all SBase attributes in columns of the table that represents the SBase, as well as the transformation of all the constraints associated to the SBase in constraints associated to the table that represents the SBase. The calls to the SBase2Stable and SpecializedBase2STable relations permit to cover recursively through all the hierarchy of bases that conforms the dimension. The SpecializedBase2Stable relation has certain similarity with the SBase2Stable relation, for that reason we are not going to define it.

In Fig. 16 we have omitted the attributes in some tables, as well as the primary keys and the foreign key to make the scheme snowflake more understandable. Be observed how the security constraints have been modelled at the logical level. To complete the case study only remains apply the AssocSF_SD2FKey, AssocSDF_SD2FKey and AssocSDF_SF2FKey relations, which enable to establish relationships between SFact and the SDimensions, between the SDegenerateFact and the SDimensions and between the SFact and the SDegenerateFact.

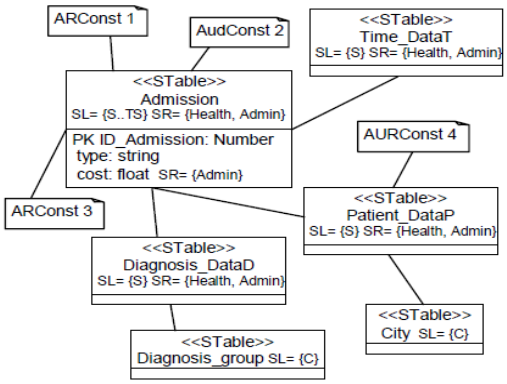


Figure 16 Snowflake schema representing a instance of the SMD PSM

In our case only proceeds to establish relations between the SFact Admission and the SDimensions, therefore we do not have SDegenerateFact. As consequence, when the AssocSF_SD2FKey relation is applied, then three foreign keys are created the Admission table. These keys enable the relationships between the Admission table with the Diagnosis_DateP, Patient_DataP and Time_DataT tables.

V. CONCLUSION

We have accomplished an important step towards completing our MDA architecture for developing secure Data Warehouses with presenting rules to transform the proposed conceptual model in [8] to logical model in this paper. The proposed model considers DW temporal and security requirements; the model is divided to two stages ETL stage and DW stage. We extend our approach by defining QVT relations in order to automatically transform the MD conceptual model into logical models that are closer to the relational implementation.

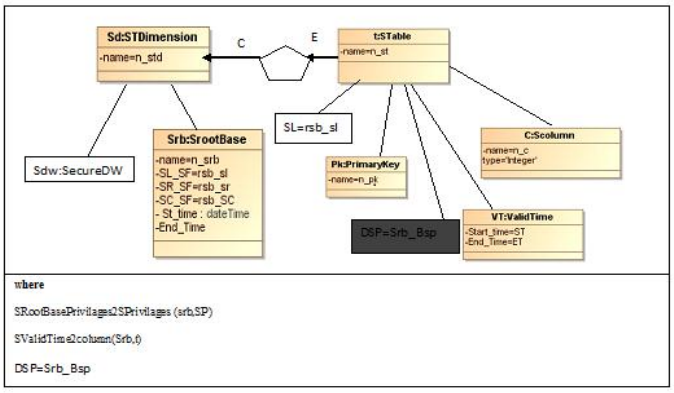


Figure 15 Slowly Changing Dimension(SCDSDimension)2STable

Our MDA architecture for developing secure DWs allows us to define security requirements and temporal issues in DW but should be extended with possibility to use the multidimensional capabilities in any commercial tools as SQL

Server Analysis Services to complete our MDA framework. Our immediate future work is to translate the security measures defined in our secure multidimensional model at conceptual model and logical model into secure multidimensional code for any commercial tool.

REFERENCES

[1] Object Management Group: MDA Guide 1.0.1, http://www.omg.org/cgi-bin/doc?omg/03-06-01,2003

[2] Object Management Group: MOF 2.0 Query/Views/Transformations, http://www.omg.org/cgi-bin/doc?ptc/2005-11-01,2005

[3] .NORBERTO MAZÓN, J.PARDILLO AND J. TRUJILLO,"APPLYING TRANSFORMATIONS TO MODEL DRIVEN DATA WAREHOUSES", DISCOVERY LECTURE, VOL 4081, PP.13-22,2006

[4] A. ABELL´O, , J. SAMOS AND F. SALTOR,"A FRAMEWORK FOR THE CLASSIFICATION AND DESCRIPTION OF MULTIDIMENSIONAL DATA MODELS". IN: DEXA.LECTURE NOTES IN COMPUTER SCIENCE, SPRINGER VOL. M2113, PP. 668–677,2001

[5] M. Golfarelli, S. Rizzi," Methodological framework for data warehouse design". In: DOLAP, ACM (1998) 3–9

[6] N. Tryfona, F. Busborg and J.G.B. Christiansen, "starER: A conceptual model for data warehouse design", **DOLAP '99:** Proceedings of the 2nd ACM international workshop on Data warehousing and OLAP, 1999

[7] S. Luj´an-Mora, J. Trujillo and I.Y.Song, "A UML profile for multidimensional modeling in data warehouses". Data& Knowledge Engineering ,Vol. 59, Issue 3, PP. 725–769, December 2006.

[8] M.S. Farhan, M.E. Marie,L.M.E Fangary and Y.K, Helmy, "An Integrated Conceptual Model for Temporal Data Warehouse Security". Computer and Information Science, Vol 4, No 4, pp.46-57, 2011

[9] A. Simitsis, P. Vassiliadis, and T. K. Sellis." Optimizing ETL processes in data warehouses". In Proc. ICDE,pp. 564–575, 2005.

[10] P. Vassiliadis, A. Simitsis, P. Georgantas, M. Terrovitis, and S. Skiadopoulos. "A generic and customizable framework for the design of ETL scenarios". Information Systems,vol.30(7),pp.492–525, 2005.

[11] A. Simitsis. "Mapping conceptual to logical models for ETL processes" In Proc. DOLAP,2005.

[12] P. Vassiliadis, A. Simitsis, and S. Skiadopoulos." Conceptual modeling for ETL processes". In Proc. DOLAP, pp. 14–21, 2002

[13] M. Bouzeghoub, F. Fabret, and M. Matulovic. "Modeling data warehouse refreshment process as a workflow application". In Proc. DMDW, 1999.

[14] D. Calvanese, G. De Giacomo, M. Lenzerini, D. Nardi, and R. Rosati. Information integration:" Conceptual modeling and reasoning support". In Proc. CoopIS, pp. 280–291, 1998.

[15] R. Kirkg¨oze, N. Katic, M. Stolda, and A. M. Tjoa. "A security concept for OLAP". In Proc. DEXA, pp 619–626, 1997.

[16] S. Jajodia and D. Wijesekera. "Securing OLAP data cubes against privacy breaches". IEEE Symposium on Security and Privacy Proceedings, pp. 161–178, 2004.

[17] T. Priebe and G. Pernul. "A pragmatic approach to conceptual modeling of OLAP security". In Proc. ER,pp. 311–324, 2000.

[18] E. Fernandez-Medina, J. Trujillo, R. Villaroel, and M. Piattini. "Extending UML for designing secure data warehouses". In Decision Support Systems, 2006.

[19] S. Rizzi, A. Abell´, J.Lechtenb¨orger, and J. Trujillo, "Research in Data Warehouse Modeling and Design: Dead or Alive? "In Proc. of DOLAP'06,2006

[20] M. Mrunalini, T.V. Suresh Kumar and K. Rajani Kanth, "Simulating Secure Data Extraction in Extraction Transformation Loading (ETL) Processes," Third UKSim European Symposium on Computer Modeling and Simulation, 2009

[21] J. Jürjens, "Secure Systems Development with UML", Springer-Verlag, USA,2004

[22] D. Basin, J. Doser, and T. Lodderstedt, 'Model driven security: from UML models to access control infrastructures', *ACM Transactions on Software Engineering and Methodology*, Vol. 15, No. 1, pp.39–91,2006

[23] T. Lodderstedt, D. Basin and J. Doser,"SecureUML: a UML-based modeling language for model-driven security', UML 2002, The Unified Modeling Language, Model Engineering, Languages Concepts, and Tools, 5th International Conference, Springer, Dresden, Germany,2002

[24] [24] B. Thuraisingham, M. Kantarcioglu and S. Iyer, 'Extended RBAC-based design and implementation for a secure data warehouse', International Journal of Business Intelligence and Data Mining (IJBIDM), Vol. 2, No. 4, pp.367–382,2007

[25] T. Priebe and G. Pernul "A pragmatic approach to conceptual modeling of OLAP security", 20th International Conference on Conceptual Modeling (ER 2001), Springer-Verlag, Yokohama, Japan, 2001

[26] E. Fernández-Medina, J. Trujillo, R. Villarroel and M. Piattini, "Developing secure data warehouses with a UML extension", Information Systems, Vol. 32, No. 6, pp.826–856,2007

[27] C. Blanco,I. García-Rodríguez de Guzmán, I., Rosado, D.G., E. Fernandez-Medina, J. Trujillo, "Applying QVT in order to implement secure data warehouses in SQL server analysis services", Journal of Research and Practice in Information Technology, Vol. 41, No. 2, pp.119–138..2009

[28] J.Norberto Mazón, J.Trujillo and J. Lechtenbörger." A Set of QVT Relations to Assure the Correctness of Data Warehouses by Using Multidimensional Normal Forms.", 25th International Conference on Conceptual Modeling, Tucson, AZ, USA, November 6-9, 2006, Proceedings. Volume 4215 of Lecture Notes in Computer Science, pages 385-398, Springer, 2006.

[29] E. Soler, , J. Trujillo, E. Fernandez-Medina and M.Piattini,"Application of QVT for the Development of Secure Data Warehouses: A case study" The Second International Conference on Availability, Reliability and Security, ARES 2007.

[30] J.Pardillo, J. Mazón, "Designing OLAP schemata for data warehouses from conceptual models with MDA", 2010

[31] R. Kimball and M. Ross, *The Data Warehousing Toolkit*, 2nd edition:John Wiley, 2002

[32] C. Blanco,I. García-Rodríguez de Guzmán and E.Fernández-Medina "Defining and Transforming Security Rules in an MDA Approach for DWs". *Int. J. Business Intelligence and Data Mining, Vol. 5, No. 2, 2010*

Web Anomaly Misuse Intrusion Detection Framework for SQL Injection Detection

Shaimaa Ezzat Salama, Mohamed I. Marie, Laila M. El-Fangary & Yehia K. Helmy
Information System Department,
Faculty of Computers and Information
Helwan University, Cairo, Egypt

***Abstract*—Databases at the background of e-commerce applications are vulnerable to SQL injection attack which is considered as one of the most dangerous web attacks. In this paper we propose a framework based on misuse and anomaly detection techniques to detect SQL injection attack. The main idea of this framework is to create a profile for legitimate database behavior extracted from applying association rules on XML file containing queries submitted from application to the database. As a second step in the detection process, the structure of the query under observation will be compared against the legitimate queries stored in the XML file thus minimizing false positive alarms.**

Keywords-SQL injection; association rule; anomaly detection; intrusion detection.

I. INTRODUCTION

Database-driven web applications have become widely deployed on the Internet, and organizations use them to provide a broad range of services to their customers. These applications, and their underlying databases, often contain confidential, or even sensitive, information, such as customer and financial records. However, as the availability of these applications has increased, there has been a corresponding increase in the number and sophistication of attacks that target them. One of the most serious types of attack against web applications is SQL injection. In fact, the Open Web Application Security Project (OWASP), an international organization of web developers, has placed SQL injection attack (SQLIA) at the top of the top ten vulnerabilities that a web application can have [1]. Similarly, software companies such as Microsoft have cited SQLIAs as one of the most critical vulnerabilities that software developers must address [2]. As the name implies, this type of attack is directed toward database layer of the web applications. Most web applications are typically constructed in a two- or three-tiered architecture as illustrated in Fig.1 [3].

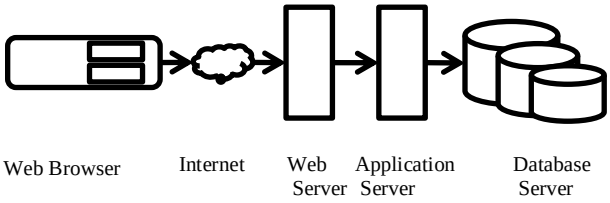


Figure 1. three-tiered architecture

SQLIA is a type of code-injection attack in which an attacker uses specially crafted inputs to trick the database into executing attacker-specified database commands. SQLIAs can give attackers direct access to the underlying databases of a web application and, with that, the power to leak, modify, or even delete information that is stored on them. The root cause of SQLIAs is insufficient input validation [4, 5]. SQLIAs occur when data provided by a user is not properly validated and is included directly in a SQL query [6]. We will provide a simple example of SQLIA to illustrate the problem.

*Select * from users where user_name='' & name & '' and password='' & pass & ''*

The previous example works well if the user supplies valid user name and password. But the problem arises when malicious user exploits the invalidated input and changes the structure of the query to achieve one or more of the different attack intents [4, 7]. The structure of the query will be altered if the user_name attribute have the following value: ' or 1=1 --. The full text of the previous query becomes:

*Select * from users where user_name='' or 1=1*

The injected code will delete the password constraint through the use of SQL comment - - and makes the condition of the query always evaluate to true.

One mechanism to defend against web attacks is to use intrusion detection systems (IDS) and especially network intrusion detection systems (NIDS). IDS use misuse or anomaly or both techniques to defend against attacks [8]. IDS that use anomaly detection technique establish a baseline of normal usage patterns, and anything that widely deviates from it gets flagged as a possible intrusion. Misuse detection technique uses specifically known patterns of unauthorized behavior to predict and detect subsequent similar attempts. These specific patterns are called signatures [8,9].

Unfortunately, NIDS are not efficient or even useful in web intrusion detection. Since many web attacks focus on applications that have no evidence on the underlying network or system activities, they are seen as normal traffic to the general NIDS and pass through them successfully [7, 10, 11].

NIDS are mostly sitting on the lower (network/transport) level of network model while web services are running on the higher (application) level as illustrated in Fig. 2 [11].

In this paper, we propose a framework that combines the two IDS techniques, misuse and anomaly detection techniques, to defend against SQLIA. The main idea of Web Anomaly Misuse Intrusion Detection (WAMID) framework is to create a profile for web application that can represent the normal behavior of application users in terms of SQL queries they submit to the database. Database logs can be used to collect these legitimate queries provided that these logs are free of intrusions. We then use an anomaly detection model based on data mining techniques to detect queries that deviates from the profile of normal behavior. The queries retrieved from database log are stored in XML file with predefined structure. We choose XML format because it is more structured than flat files, more flexible than matrices, simpler and consume less storage than databases.

Association rules will be applied to this XML file to retrieve relation between each table in the query with each condition in the selection part. These rules represent the profile of normal behavior and any deviation from this profile will be considered attack. In order to better detect SQLIA and to minimize false positive alerts, WAMID framework as a second step uses misuse technique to detect any change in the structure of the query. Malicious users sometimes don't change the selection clause but add another SQL statement or add specific keywords to the initial query to check the vulnerability of the site to SQLIA or to perform inference attack. Such types of attack are detected in the second step of the detection process. By comparing the structure of the query under test with the corresponding queries in the XML file the previous malicious actions will be detected.

The rest of the paper will be organized as follows: in section II discusses previous work, section III

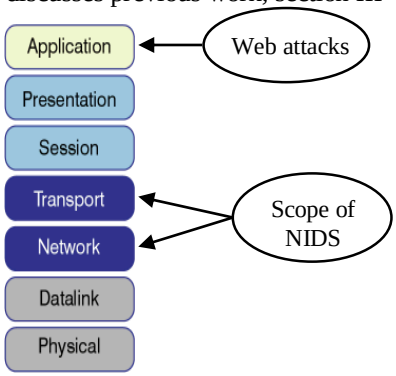


Figure 2. Scope of NIDS

provides a detailed description about the framework and its components. Anomaly and misuse algorithms and a working example will be presented in section IV. Section V concludes the paper and outlines future work.

II. Literature Review

Different researches and approaches have been presented to address the problem of web attacks against databases. Considering SQLIA as top most dangerous attacks, as stated in section I, there has been intense research in detection and prevention mechanisms against this attack [4, 5, 12]. We can classify these approaches into two broad categories: a) one approach is trying to detect SQLIA through checking anomalous SQL query structure. b) another approach uses data dependencies among data items which are less likely to change for identifying malicious database activities. In either of two categories, different researchers take advantage of the benefit of integrating data mining with database intrusion detection in order to minimize false positive alerts, minimizing human intervention and better detect attacks [13]. Moreover, Different intrusion detection techniques are used either separately or together. Different work used misuse technique others used anomaly or mixes the two techniques.

Under the first category and without using data mining technique, Lee et al. in [10] and Low et al. in [14] developed a framework based on fingerprinting transactions for detecting malicious transactions. They explored the various issues that arise in the collation, representation and summarization of this potentially huge set of legitimate transaction fingerprints. Another work that applies anomaly detection technique to identify anomalous database application behavior is presented by Valeur et al. in [15]. It builds a number of different statistical query models using a set of typical application queries, and then intercepts the new queries submitted to the database to check for anomalous behavior.

A general framework for detecting malicious database transaction patterns using data mining was proposed by Bertino et al. in [16, 17] to mine database logs to form user profiles that can model normal behaviors and identify anomalous transactions in databases with role based access control mechanisms. The system is able to identify intruders by detecting behaviors that differ from the normal behavior of a role in a database. Kamra et al. in [18] illustrated an enhanced model that can also identify intruders in databases where there are no roles associated with each user. It employs clustering techniques to form concise profiles representing normal user behaviors for identifying suspicious database activities. Another approach that checks for the structure of the query to detect malicious database behavior is the work of Bertino et al. in [19]. They proposed a framework based on anomaly detection technique and association rule mining to identify the query that deviates from normal database application behavior.

The problem with this framework is that it produces a lot of rules and represents the queries in very huge matrices which may affect tremendously on the performance of rule extraction. Misuse detection technique have been used by Bandhakavi et al. in [20] to detect SQLIA by discovering the intent of a query dynamically and then comparing the structure of the identified query with normal queries based on the user input with the discovered intent. The problem with this approach is that it has to access the source code of the application and make some modifications to the java virtual machine.

Halfond et al. in [21] developed a technique that uses a model-based approach to detect illegal queries before they are executed on the database. In its static part, the technique uses program analysis to automatically build a model of the legitimate queries that could be generated by the application. In its dynamic part, the technique uses runtime monitoring to inspect the dynamically-generated queries and check them against the statically-built model. The system WASP proposed by Wiliam et al. in [22] tries to prevent SQL Injection Attacks

by a method called positive tainting. In positive tainting, the trusted part of the query (static string) is not considered for execution and masked as tainted, while all other inputs are considered. The difficulty in this case is the propagation of taints in a query across function calls especially for the user defined functions which call some other external functions leading to the execution of a tainted query. Different other researches followed the same approach in detection of anomalous SQL query structure in [23, 24].

Researches that belong to the second category of detection which depends on data dependencies are [25, 26, 27, 28]. The work that is based on mining sequential data access patterns for database intrusion detection was proposed by Hu et al. in [25, 26]. Transactions that do not comply with rules generated from read and write sequence sets are identified as malicious transactions. Srivastava et al. offered a weighted sequence mining approach [27] for detecting database attacks. The advantage of the work presented by YiHu et al. in [28] is the automatic discovery and use of essential data dependencies, namely, multi-dimensional and multi-level data dependencies, for identifying anomalous database transactions.

The contribution of this paper is a framework that combines anomaly and misuse detection technique in order to better detect SQLIA. This framework uses association rules with anomaly technique to build the normal behavior of application users and detecting anomalous queries. Moreover, misuse technique is used to check the structure of the query to detect any malicious actions that cannot be detected using anomaly detection technique.

III. THEORETICAL FRAMEWORK

WAMID framework is a database intrusion detection that aims to detect SQLIA at real-time, before queries execution at the database. This is why this framework should run at the database or application server depending on the architecture of web application as depicted in fig. 1. In order to detect all possible attempts of SQL injection, WAMID framework combines the two detection techniques: anomaly and misuse. It depends in the detection of SQLIA on determining the malicious changes that occurred in the SQL query structure. The key idea of our framework is as follows. We build a repository containing set of legitimate queries submitted from the application user to the database. This repository is a set of training records. We then use an anomaly detection approach based on data mining technique to build a profile of normal application behavior and indicate queries that deviates from this normal behavior.

In a second step in the detection process, the framework checks for the presence of dangerous keywords in the query if the latter passes the test of anomaly detection step. We need this step because sometimes the intent of the attacker is to identify the security holes in the site or to infer the structure of the database through the error message returned from the application and this type of SQLIA is called inference [4, 29]. This type of attack cannot be detected through anomaly technique because it doesn't require change in the conditions of the original query but it will be discovered if the structure of the query is compared against its corresponding query in the repository file.

Based on what previously stated we learn that the framework act in two phases: training phase and detection phase. In the training phase the repository file will be created and normal behavior of the application is built. In the detection phase, the framework uses the anomaly and misuse techniques to discover any SQLIA. In the following subsections we will provide a detailed explanation of the framework, its components and how it works.

A. Training Phase

During the training phase the training records are collected from the queries the application send to the database. The source for obtaining these query traces is the database log provided that the latter is free of intrusions. The training phase flow is illustrated in Fig. 3. The challenge here is that to efficiently encode these queries in order to extract useful features from them and accordingly build the application fingerprint. Unlike approach provided in [19], we choose to encode the queries in XML file. The encoding scheme provided by Bertino et al. in [19] result in a large, dense, sparse matrices which may effect on the mining algorithm. XML is more structured than flat files, is supported by query tools like XQuery and XPath to extract data [30]. It is simpler and consumes less space than relational databases and more flexible than matrices.

It is important to identify accurately the structure of the XML file that will represent the features extracted from the query that will contribute in building the application fingerprint. Consider the following query:

Select SSN, last_name from employee where first_name='Suzan' and salary>5000

The encoding scheme of the previous query in XML file is illustrated in Fig. 4. The main advantage of XML format is that nodes may be duplicated upon need. For example the number of project_attribute" node may differ from one "Query" node to another depending on the query itself. This is why it is more suitable to store queries than databases while maintaining flexibility and simplicity.

The XML file illustrated in fig. 4 stores the projection attributes, the from clause and the predicate clause in a more detailed way. It is not important to identify the value of the integer or string literal it is important to determine that there is an integer or string literal or there is another attribute in the right hand side. Another file that should be created during the training phase is the signature file that will be used during the misuse detection phase. As stated before this file contains suspicious keywords that may be considered a sign of SQLIA.

Keywords like for example single quote, semicolon, double dash, union, exec, order by and their hexadecimal representation in order to prevent the different evasion techniques [31]. The important step in the training phase is to build the profile representing the application normal behavior. We will apply association rules [32] on the XML file to extract rules that represent the normal behavior of application users. Different approaches have been proposed to apply association rules on XML data. We direct the reader to [33-35] for an in-depth survey of these approaches. The rules extracted represent

relationship between each table in the query with each predicate in the selection clause.

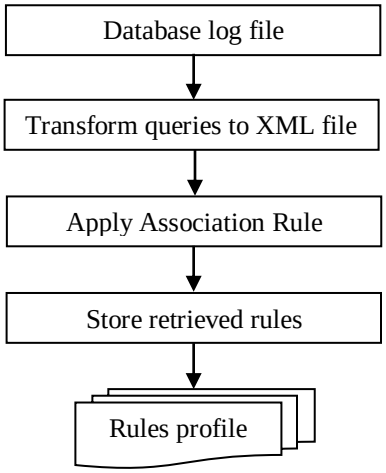


Figure 3. training phase flow

```
<Queries>
<Query id=1>
<command> select </command>
<project_attribute> SSN </project_attribute>
<project_attribute> last_name </project_attribute>
<From> employee </From>
<LHS_condition> first_name </LHS_condition>
<RHS_condition> string Literal </RHS_condition>
<logical_operator> and </logical_operator>
<LHS_condition> salary </LHS_condition>
<RHS_condition> Integer Literal </RHS_condition>
</Query>
</Queries>
```

Figure 4. representation of query in XML file

This is based on an observation that the static part of the query is the projection attribute and the part that is constructed during execution is the selection part [19]. We here add another item to the static part which are the tables in the from clause. We try to make relation between the static part and the dynamic part and extract rule with support and confidence of such relation. Any query that will not match rules extracted and stored in the rules profile will be considered attack. More details about how the rules are extracted are provided in the following subsection.

B. Anomaly Detection Phase

In the previous subsection, we illustrated how the benign queries are collected and captured in XML file in a form enabling the framework from creating the database behavior profile. We apply association rules on the XML file containing legitimate queries and extract rules that can describe the normal behavior of application users. The idea behind building the profile rule is to apply one of association rules algorithms on previously created XML file to extract relation between each table in the query with each selection attribute excluding the literals. Thus the rules extracted have the following format:

$$\text{From} \longrightarrow \text{LHS}$$

$$\text{From} \longrightarrow \text{RHS}$$

Recall the example of employee first name and salary so the rules extracted from this query are:

$$\text{employee} \longrightarrow \text{first\_name}$$

$$\text{employee} \longrightarrow \text{salary}$$

The rules that exceed the minimum support and confidence will be stored in rules profile. These rules represent the profile of how the application behaves normally. Fig. 4 illustrates the flow of detection phase of the framework in general including the anomaly technique. In a typical database application, the input supplied by the user construct the where clause of the query. Meanwhile, the projection clause and the from clause remain static at the run time. So we create a relation between the static and the dynamic part of the query and any change in the where clause by attackers that cannot be derived from the rules profile will be announced as SQLIA. We decided to choose the tables in the from clause from the static part of the query instead of the projection attributes because the former is more general and contain the latter and thus generating less rules and make it easier in comparison. Lets return to our query in the previous subsection and change it a little bit: select SSN, last_name from employee where first_name=' "& fname &" ' and salary> " & empsal. If the attacker needs to retrieve all values from employee table then the following code will be injected to form this new query:

Select SSN, last_name from employee where first_name='' or 1=1 - -

Before executing this query, rules should be extracted first and compared to the rules in the rules profile. The relation between tables and attributes will be compared against rules stored in the profile rules file. The two relations under test from the previous example are:

$$\text{employee} \longrightarrow \text{first\_name}$$

$$\text{employee} \longrightarrow 1$$

The first relation exists in the rules profile but no such rule match the second relation. So the query is announced as anomaly query.

C. Misuse Detection Phase

In a second step in the detection process and after the anomaly detection phase, comes the role of misuse detection. The need to this step comes from the fact that SQLIA doesn't only change the conditions in the query but it also may provide information about the database schema or check the vulnerability of the application to SQL injection. This is done through adding to the query some keywords that may change the behavior of the query or return information about the database through database errors without changing the predicates of the query. In such case, the anomaly detection phase will not be able to discover such attack. For example consider the following query:

*Select * from employee where SSN=10*

If the attacker just adds a single quote at the end of the query, this will result in error message that may inform the attacker that the site is vulnerable to SQLIA. Another example of attack is just adding the keyword "order by" to the query without changing the selection attributes like:

*Select * from employee where SSN=10 order by 1*

Trying to execute this query several times will give attacker information about the number of attributes in the table. This is why this step is needed in the detection process. Moreover, the framework doesn't announce the query as anomaly just by finding these keywords in the query because it may be part of the legitimate query itself resulting in false positive alarm. This is why the framework checks for the structure of the query under test with the corresponding query stored in XML file. The detection phase flow of the framework in Fig. 4 illustrates this process. These suspicious keywords are stored in file called "forbidden keywords". This file contains SQL keywords like single quote,

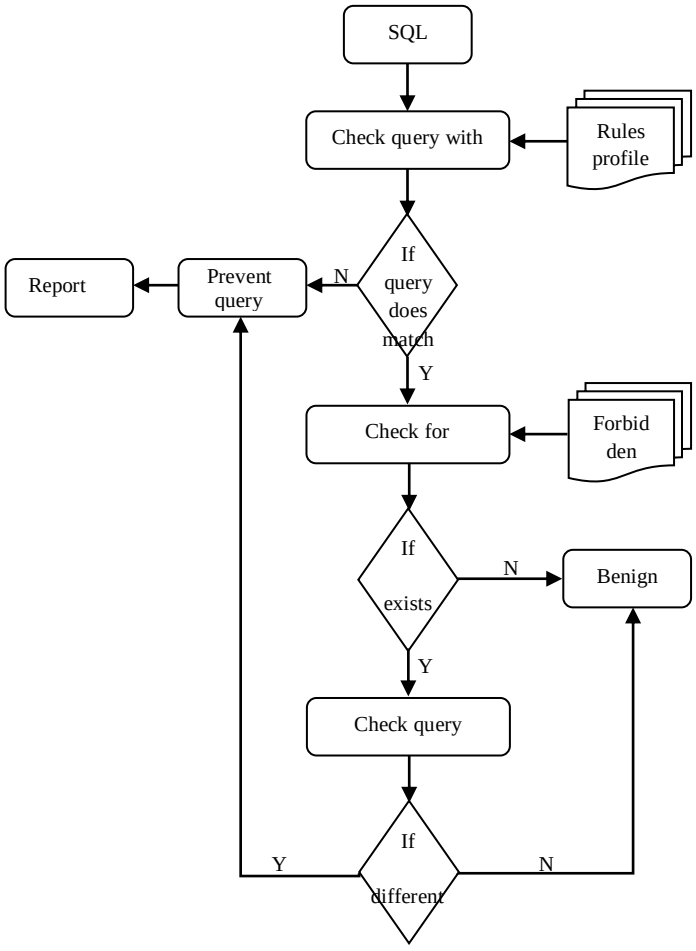


Figure 5. anomaly misuse detection flow phase

semicolon, union, order by, exec and their hexadecimal representation to avoid the different evasion techniques. After confirming the existence of one or more of these keywords, we use XQuery to retrieve queries from XML file with the same projection attributes and same from clause. Then comparison is done between query under test and the queries retrieved by XQuery from XML file. If there is no match then the query is announced anomaly.

IV. Algorithm And Working Example

In this section we present algorithms for anomaly and misuse detection. In addition, we provide a working example illustrating how the WAMID framework performs the detection.

A. Anomaly detection algorithm

Algorithm anomaly_detection( )
Input: rules profile, query under test
Output: True if query is intrusion, false otherwise
Begin

Extract relation between tables and selection attributes
Store extracted relations in query_relation
/* query_relation is array to store relations*/
For each relation r in query_relation
If (r is found in rule profile)
Score=score+1
If score=length (query_relation)
Return false
Else
return true
End

B. Misuse detection algorithm

Algorithm misuse_detection( )
Input: forbidden keywords file, query under test, XML file
Output: True if query is intrusion, false otherwise
Begin
For each keywords k in forbidden keywords
If k not exists in query
Return false
Else
Use XQuery language to extract relevant queries from XML file
If query structure doesn't match any retrieved queries
Return True
Else
Return false
End

C. Working example

In order to provide better understanding of the anomaly and misuse detection in WAMID framework, we provide in this subsection example of the flow of detection either anomaly or misuse in this framework. The following represents example of queries submitted from application to database:

- Select product_name, description from product where product_id=?
- Select product_name, description from product where salary<?
- Select * from product where product_name=? order by product_name

- Select product_name, description from product where salary<? and category_id=?

The representation of the previous queries in XML file is illustrated in Fig. 5.

```
<Queries>
<Query id=1>
<command> select </command>
<project_attribute > product_name </project_attribute>
<project_attribute > description      </project_attribute>
<from> product </from>
<LHS>product_id </LHS>
<RHS> Integer_literal </RHS>
</Query>
<Query id=2>
<command> select </command>
 <project_attribute > product_name </project_attribute>
<project_attribute > description      </project_attribute>
<from> product </from>
<LHS> salary </LHS>
<RHS> Integer_literal </RHS>
</Query>
<Query id=3>
<command> select </command>
<project_attribute > * </project_attribute>
<from> product </from>
<LHS> product_name </LHS>
<RHS> string_literal </RHS>
<order by> product_name</order by>
</Query>
<Query id=4>
<command> select </command>
<project_attribute > product_name </project_attribute>
<project_attribute > description      </project_attribute>
<from> product </from>
<LHS> salary </LHS>
<RHS> Integer_literal </RHS>
<logical_operator> and </logical_operator>
<LHS> category_id </LHS>
<RHS> Integer_literal </RHS>
</Query>
```

Figure 6. XML file representing queries

After applying association rule algorithm like for example Apriori on this XML file, the resulting rules will stored in rules profile file in Fig. 6.

In the following we will provide sample of malicious and legitimate queries.

- Select product_name, description from product where product_id=5’

The first step in the framework is to identify relation between tables and selection attributes in the query.

Product ⟶ product_id

Second, the framework searches in the rules profile for this relation. It already exists. But this is not the end of the detection flow. The second step is to check for suspicious keywords in the query. The query already contains one of the suspicious keywords which is single quote.

So XQuery language is used to extract queries from the XML file with same projection attributes and same from clause. By comparing the structure of the query under test and query returned from the XML file we will find that query shouldn’t contain the single quote and thus it is announced as anomaly.

- Select product_name, description from product where product_id=1 or 1=1- -

The value 1 for the product_id may be right or maybe wrong anyway we have here an injected code to retrieve data of all products. First we extract relations.

Product ⟶ product_id relation 1
Product ⟶ 1 relation 2

By searching in the rules profile we find a rule for the first relation but no rule for the second relation so the query is announced immediately anomaly.

- Select product_name, description from product where product_id=1 order by 1- -

As we previously stated there is a rule matching the relation 1 in the previous example.

By examining the query against the forbidden keywords file we find two keywords: order by and double dash. By examining the original query in the XML file we find that this query is anomaly because it doesn’t contain order by or double dash.

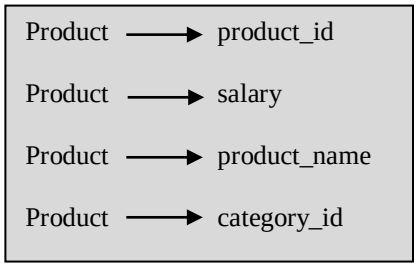


Figure 7. Extracted rules from XML file

- Select * from product where product_name=’food’ order by product_name

The extracted relation from this query is:

Product ⟶ product_name

This relation exists in the rules profile. And also one of the forbidden keywords exists so the structure of the query should be examined. After examining the structure of the query the framework identifies that the query is legitimate.

V. Conclusion And Future Work

Database intrusion is a major threat to any organization storing valuable and confidential data in databases. This is increasingly more so as the number of database servers connected to the Internet increases rapidly. Existing network-based and host-based intrusion detection systems are not sufficient for detecting database intrusions. We have introduced a framework based on anomaly and misuse detection for discovering SQLIA. We have presented a new encoding technique for SQL queries in XML file in a way enabling the extraction of normal behavior of database application. We then used data mining technique for fingerprinting SQL statements and use them to identify SQLIA. This set of fingerprints is then used to match incoming

database transactions. If the set of fingerprints in the legitimate set is complete, any incoming transaction whose fingerprint does not match any of those in the legitimate set is very likely to be an intrusion. A second step in the framework is the misuse technique in which XQuery is used to match the incoming query with queries stored in XML file after ensuring that one or more of the suspicious keywords exist in the query.

We plan to perform experiments to apply this framework to identify its performance in detecting attacks and include comparisons to other approaches. This work may be extended to include detection against other attacks like cross site scripting.

REFERENCES

[1] [1] http://www.owasp.org/index.php, OWASP Top 10-2010 document

[2] M. Howard and D. LeBlanc, "Writing Secure Code", Microsoft Press, 2002

[3] Amit Kumar Pandey, "SECURING WEB APPLICATIONS FROM APPLICATION-LEVEL ATTACK", master thesis, 2007

[4] W.G.Halfond, J.Viegas, and A.Orso, "A classification of SQL-Injection Attacks and Countermeasures", in proceeding of the International Symposium on Secure Software Engineering (ISSSE), 2006

[5] Kindy, D.A.; Pathan, A.K, "A survey on SQL injection: Vulnerabilities, attacks, and prevention techniques", in proceedings of IEEE 15th International Symposium on Consumer Electronics (ISCE), 2011

[6] G. Wassermann and Z. Su, "An Analysis Framework for Security in Web Applications", In Proceedings of the FSE Workshop on Specification and Verification of Component-Based Systems (SAVCBS 2004), pages 70–78, 2004.

[7] San-Tsai Sun, Ting Han Wei and Stephen Liu, "Classification of SQL Injection Attacks", University of British Columbia : Sheung Lau Electrical and Computer Engineering, 2007

[8] S.Axelsson, "Intrusion detection systems: A survey and taxonomy", Technical Report, Chalmers Univ., 2000

[9] Marhusin, M.F.; Cornforth, D.; Larkin, H., "An overview of recent advances in intrusion detection", in proceeding of IEEE 8th International conference on computer and information technology CIT, 2008

[10] Lee, S. Y., Low, W. L., and Wong, P. y.: Learning Fingerprints for a Database Intrusion Detection System. In the Proceedings of the 7th European Symposium on Research in Computer Security, 2002

[11] C.J. Ezeife, J. Dong, A.K. Aggarwal, "SensorWebIDS: A Web Mining Intrusion Detection System", International Journal of Web Information Systems, volume 4, pp. 97-120, 2007

[12] N. Khochare, S. Chalurkar ,S. Kakade, B.B. Meshramm, "Survey on SQL Injection attacks and their countermeasures", International Journal of Computational Engineering & Management (IJCEM), Vol. 14, October 2011

[13] S. F. Yusufovna., "Integrating Intrusion Detection System and Data Mining", International Symposium on Ubiquitous Multimedia Computing, 2008

[14] Low, W. L., Lee, S. Y., Teoh, P., "DIDAFIT: Detecting Intrusions in Databases Through Fingerprinting Transactions", in Proceedings of the 4th International Conference on Enterprise Information Systems (ICEIS), 2002

[15] F. Valeur, D. Mutz, and G.Vigna, "A learning-based approach to the detection of sql injection attacks", in proceedings of the conference on detection of intrusions and Malware and vulnerability assessment (DIMVA), 2005

[16] Bertino, E., Kamra, A, Terzi, E., and Vakali, A, "Intrusion detection in RBAC-administered databases", in the Proceedings of the 21st Annual Computer Security Applications Conference, 2005

[17] Kamra A, Bertino, E., and Lebanon, G.,"Mechanisms for Database Intrusion Detection and Response", in the Proceedings of the 2nd SIGMOD PhD Workshop on Innovative Database Research, 2008

[18] Kamra A, Terzi E., and Bertino, E.,"Detecting anomalous access patterns in relational databases", the VLDB Journal VoU7, No. 5, pp. 1063-1077, 2009

[19] Bertino, E., Kamra, A, and Early, J., "Profiling Database Application to Detect SQL Injection Attacks", In the Proceedings of 2007 IEEE International Performance, Computing, and Communications Conference, 2007

[20] Bandhakavi, S., Bisht, P., Madhusudan, P., and Venkatakrishnan V., "CANDID: Preventing sql injection attacks using dynamic candidate evaluations", in the Proceedings of the 14th ACM Conference on Computer and Communications Security, 2007

[21] Halfond, W. G. and Orso, A , "AMNESIA: Analysis and Monitoring for Neutralizing SQL-Injection Attacks", in Proceedings of the 20th IEEE/ACM international Conference on Automated Software Engineering, 2005

[22] William G.J. Halfond, Alessandro Orso, and Panagiotis Manolios, "WASP: Protecting Web Applications Using Positive Tainting and Syntax-Aware Evaluation", IEEE Transactions on Software Engineering, Vol. 34, No. 1, pp 65-81, 2008

[23] Buehrer, G., Weide, B. w., and Sivilotti, P. A, "Using Parse Tree Validation to Prevent SQL Injection Attacks", in Proceedings of the 5th international Workshop on Software Engineering and Middleware, 2005

[24] Liu, A, Yuan, Y., Wijesekera, D., and Stavrou, A, "SQLProb:A Proxy-based Architecture towards Preventing SQL Injection Attacks", in Proceedings of the 2009 ACM Symposium on Applied Computing, 2009

[25] Hu, Y., and Panda, B., "A Data Mining Approach for Database Intrusion Detection", In Proceedings of the 19th ACM Symposium on Applied Computing, Nicosia, Cyprus ,2004

[26] Hu, Y., and Panda, B., "Design and Analysis of Techniques for Detection of Malicious Activities in Database Systems", Journal of Network and Systems Management, Vol. 13, NO. 3,2005

[27] Srivastava, A, Sural S., and Majumdar, AK., "Database Intrusion Detection Using Weighted Sequence Mining", Journal of Computers, vol.1, no. 4 ,2006

[28] Yi Hu; Campan, A.; Walden, J.; Vorobyeva, I.; Shelton, J, "An effective log mining approach for database intrusion detection", in proceedings of IEEE international conference on systems man and cybernetics (SMC), 2010

[29] David Litchfield, "Data-mining with SQL Injection and Inference",An NGSSoftware Insight Security Research, September 2005

[30] World Wide Web Consortium. XQuery 1.0: An XML Query Language (W3C Working Draft). http://www.w3.org/TR/2002/WDxquery-20020816, Aug. 2002.

[31] O. Maor and A. Shulman, "SQL Injection Signatures Evasion", White paper, Imperva, April 2004. http://www.imperva.com/application defense center/white papers/sql injection signatures evasion.html

[32] Han J., Kamber M., "Data Mining: Concepts and Techniques", Maurgan Kaufmann, 2nd edition, 2006

[33] Jacky W.W.Wan, Gillian Dobbie, "Mining Association Rules from XML Data using XQuery", in proceeding of ACM 2nd workshop on Australasian information security, Data Mining and Web Intelligence, and Software Internationalization, 2004

[34] Qin Ding, "Data Mining on XML Data", in Encyclopedia of Data Warehousing and Mining, 2nd edition, Vol. 1, ed. John Wang, IGI Global, 2008, pp. 506-510

[35] Qin Ding and Gnanasekaran Sundarraj, "Mining Association Rules from XML Data", in Data Mining and Knowledge Discovery Technologies, ed. David Taniar, IGI Global, 2008. pp. 59-71

An Improved Grunwald-Letnikov Fractional Differential Mask for Image Texture Enhancement

Vishwadeep Garg
Research Scholar, Electronics and Communication Engineering Department
Thapar University
Patiala, India

Kulbir Singh
Associate Professor, Electronics and Communication Engineering Department
Thapar University
Patiala, India

***Abstract*—Texture plays an important role in identification of objects or regions of interest in an image. In order to enhance this textural information and overcome the limitations of the classical derivative operators a two-dimensional fractional differential operator is discussed, which is an improved version of the Grunwald-Letnikov (G-L) based fractional differential operator. A two dimensional-isotropic gradient operator mask based on G-L fractional differential is constructed. This nonlinear filter mask is implemented on various texture enriched digital images and enhancement of features of image is controlled by varying the intensity factor. In order to analyze the enhancement quantitatively, information entropy and average gradient are the parameters used. The results show that with improved version of Grunwald-Letnikov, fractional differential operator information entropy of image is improved by 0.5.**

***Keywords*—texture enhancement; fractional differential; information entropy; average gradient.**

I. INTRODUCTION

Texture is an innate property of virtually all surfaces which contain important information about structural arrangement of surfaces and their relation to surrounding environment [1]. Since, the textural properties of image carry useful information for discrimination purpose hence enhancement of these textural features is an important concern. Contrast enhancement and image sharpening are known general techniques that can enhance subtle texture of an image. The contrast enhancement techniques are discussed in [2]. The classical image sharpening techniques are based on integral derivative operators like Sobel, Prewitts and Laplacian. Recently, many methods like Laplacian Pyramid, Curvelet Transform and Wavelets are introduced in [3]-[6] in order to enhance the texture information. In recent years with phenomenal research in the field of fractional calculus, it finds an important place in the field of signal processing and digital image processing. The concept of fractional calculus came into existence in 1695 with discussion between Leibniz and L'Hospital. In relation to fractional calculus, three popular definitions are known till date which is Grunwald-Letnikov (G-L), Riemann-Liouville (R-L) and Caputo [7]-[9]. Of these G-L and R-L are the popular definitions used in digital image processing. G-L based differential operator is adopted by many researchers and scholars in [10]-[12]. Recent research has been made to construct R-L based fractional differential operator presented in [13]-[14]. Although both G-L and R-L based differential operator overcomes the deficiencies of the classical gradient operators yet they suffer from some distortion. In this paper, an improved G-L fractional differential operator is presented to process digital image.

II. FRACTIONAL DIFFERENTIATION

In this section, theoretical background of Grunwald-Letnikov (G-L) based fractional differential is considered. G-L definition of fractional calculus arises from classical definition of integral differentiation by generalizing differential order from integer to fraction. Assuming $\forall\ v \in R$ (R represents real set and, [v] is its integral part) and signal $s(t) \in [a, t]$, $a < t$, $a \in R$, $t \in R$ has m ($m \in Z$, Z represents integer set) order continuous differentiation. When $v > 0$, so v-order fractional

$$D_t^v s(t) = \lim_{h\to 0} s_h^v(t) = \lim_{h\to 0} h^{-v} \sum_{m=0}^{n-1} \binom{-v}{m} s(t-mh) \tag{1}$$

where

$$\binom{-v}{m} = \frac{(-v)(-v+1)\cdots(-v+m-1)}{m!} \tag{2}$$

Thus,

$$D_t^v = \frac{h^{-v}}{\Gamma(-v)} \sum_{m=0}^{n-1} \frac{\Gamma(m-v)}{\Gamma(m+1)} s(t-mh) \tag{3}$$

Taking the duration of signal h = 1, the dispersion expression of one-dimensional signals fractional differential is expressed as

$$\begin{aligned} \frac{d^v s(t)}{dt^v} \approx\ & s(t) + (-v)s(t-1) + \frac{(-v)(-v+1)}{2} s(t-2) \\ & + \frac{(-v)(-v+1)(-v+2)}{6} s(t-3) \cdots \\ & + \frac{\Gamma(n-v)}{\Gamma(-v)\Gamma(n+1)} s(t-n) \end{aligned} \tag{4}$$

From except the first term the other n -1 nonzero coefficients are functions with respect to fractional order v. It can be observed that sum of n nonzero coefficient is nonzero, which is the significant difference between integral and fractional differential.

III. An Improved G-L Fractional Differential

In order to make fractional differential operator more precise (3) can be rewritten as [15]

$$D_t^v=\frac{h^{-v}}{\Gamma(-v)}\sum_{m=0}^{n-1}\frac{\Gamma(m-v)}{\Gamma(m+1)}\ s\left(t+\frac{vh}{2}-mh\right) \quad (5)$$

Comparing (4) and (5), (5) have introduced the signal values of s(t) on nonnodes besides supposing v = 0, ±2, ±4,....., thus considering the three nodes $s(t + h - mh)$, $s(t - mh)$ and $s(t - h - mh)$ and with 3-point Lagrange interpolation expression it implies

$$\begin{aligned} s(\xi)\approx & \frac{(\xi - t + mh)(\xi - t - h + mh)}{2h^2}s(t + h - mh) \\ & -\frac{(\xi - t - h + mh)(\xi - t + h + mh)}{h^2}s(t - mh) \\ & +\frac{(\xi - t - h + mh)(\xi - t + mh)}{2h^2}s(t - h - mh) \end{aligned} \quad (6)$$

Assuming $\xi=t+(vh/2)-mh$ and doing fractional interpolation (6) can be written as

$$\begin{aligned} s\left(t+\frac{vh}{2}-mh\right)\cong & \left(\frac{v}{4}+\frac{v^2}{8}\right)s(t + h - mh) \\ & +\left(1-\frac{v^2}{4}\right)s(t - mh) \\ & +\left(\frac{v^2}{8}-\frac{v}{4}\right)s(t - h - mh) \end{aligned} \quad (7)$$

From (5) and (7)

$$\begin{aligned} \frac{\partial^v}{\partial t}s(t)= & \frac{h^{-v}}{\Gamma(-v)}\sum_{m=0}^{n-1}\frac{\Gamma(m - v)}{\Gamma(m + 1)} \\ & \times\left[s_m+\frac{v}{4}(s_{m-1} - s_{m+1})+\frac{v^2}{8}(s_{m-1} - 2s_m + s_{m+1})\right] \end{aligned} \quad (8)$$

where $s_m=s(t - mh)$, $s_{m-1}=s(t + h - mh)$ and $s_{m+1}=s(t - h - mh)$

This is an expression of an improved G-L Fractional Differential [15]

IV. Design Of An Improved G-L Fractional Differential

Image processing with a nonlinear filter consists of moving the filter mask from point to point in an image. This linear filtering operation is given by expression [14]

$$g(x, y)=\sum_{s=-a}^{a}\sum_{t=-b}^{b}w(s, t)f(x + s, y + t) \quad (9)$$

where f(x, y) is a value of pixel and w(s, t) is a value of mask. Considering the gradient direction, the mask is designed into an m × m-size matrix T which has m layers (m is odd natural number).

There are 8 directions of T, which are 0, π/8, π/4, 3π/8, π/2, 5π/8, 3π/4 and 7π/8, respectively. From (8) it is concluded that

$$T_i=\frac{1}{\Gamma(-v)}\left[\begin{aligned} & \frac{\Gamma(i - v + 1)}{(i+1)!}+\frac{\Gamma(i - v)}{i!}\left(1-\frac{v^2}{4}\right) \\ & +\frac{\Gamma(i - v - 1)}{(i-1)!}\left(-\frac{v}{4}+\frac{v^2}{8}\right)\end{aligned}\right] \quad (10)$$

where T_i is the value of i^{th} layer of mask, $v \in R^+$

Especially, in order to make the sum of T equal to 0, it has [14]

$$T_0= -1*\sum_{i=1}^{n}8*i*T_i \quad (11)$$

Clearly, the result of convoluting with T is the sharpening edges of the image.

For the purpose of texture enhancement, the sharpening edges must add to the value of original pixel

So T must change into R

$$\begin{cases} R_i=\gamma T_i & ;(i > 0) \\ R_0=(1+\gamma)T_0 & ; (i = 0) \end{cases} \quad (12)$$

where γ is the intensity factor.

When $v \in (0, 1)$ from (11) and (12)

$$\begin{cases} R_i=\gamma T_i & ;(i > 0) \\ R_0=1-\sum_{i=1}^{n}8*i*R_i & ; (i = 0) \end{cases} \quad (13)$$

The mask of an improved G-L fractional differential is shown in Fig. 1(a)

R1	R1	R1
R1	R0	R1
R1	R1	R1

(a)

-0.2474	-0.2474	-0.2474
-0.2474	2.9792	-0.2474
-0.2474	-0.2474	-0.2474

(b)

Figure 1. (a) A 3 X 3 size mask of an improved G-L Fractional Differential. (b) A 3 X 3 size filter mask at v = 0.5 and intensity factor γ = 1

It is seen from Fig. 1(b) that sum of coefficients of mask is not equal to zero which is a prominent difference between fractional and integral differential.

V. Simulated Results And Analysis

The Fractional Differential mask discussed above is implemented on images, and image texture enhancement is analyzed by varying mask parameters.

A. Comparative Analysis with Other Fractional Differential Masks

The intensity factor of an image is kept constant at $\gamma = 1$ and the fractional differential order (v) is varied for $0 < v < 1$. The simulated results through Grunwald-Letnikov (G-L), Riemann-Liouville (R-L) and Improved G-L Fractional Differential (FD) Masks are presented in Fig. 2.

From Fig. 2, it is seen that the texture of an image gets enhanced by changing the fractional differential order. For G-L FD mask, the texture of image first increases with increase in fractional differential order but exceeding the fractional differential order beyond some value results in distortion of image. With R-L FD mask, there is slight enhancement in texture of image with an increase in fractional differential order, but the degree of enhancement decreases after some point with the further increase in fractional differential order. Processing of image with Improved G-L FD mask results in enhancement of texture of image with an increase in fractional order but as compared to G-L FD mask the distortion is less with the further increase in fractional differential order.

For quantity analysis, the FD masks are implemented on different images and its effect on information entropy and average gradient of image are shown in Table I. The average gradient reflects the clarity of the image. It can be used to measure the spatial resolution of the image, i.e., a larger average gradient means a higher spatial resolution. The information entropy shows the average information included in the image and reflects the detail information of the image.

Carefully observing Table I it is seen that the information entropy and average gradient of the image first increase with the fractional differential order thus enhancing the textural information of the image. However, after attaining a certain maximum value the information entropy starts decreasing results in loss of textural information and hence distortion of image. The R-L FD Mask provides slight improvement of approximately 0.2 in information entropy with an increase in fractional differential order but if the fractional order is exceeded beyond 0.4, the information entropy starts decreasing and at $v = 1$ the entropy reaches its lowest value. Image processing through G-L FD Mask increases the information entropy of image by 0.5 and attains a maximum value at $v = 0.4$ thus enhancing texture, however, increasing the fractional order beyond 0.4 results in decrease in information entropy and hence distortion of image.

However, with the improved G-L FD Mask the maximum information entropy achieved at fractional differential order $v = 0.4$ is more than both G-L, and R-L FD Masks thus providing more textural information. Also the decrease in information entropy through an improved G-L FD Mask with the further increase in fractional order beyond 0.4 is less than other FD Masks resulting in less distortion of image.

B. The Intensity Factor

From (13) the value of fractional differential order is kept constant at $v = 0.2$ and intensity factor γ is varied for $1 < \gamma < 2$. The intensity factor γ is also the scale for controlling the degree of enhancement. The simulated results are presented in Fig. 3.

Observing the Fig. 3 it is seen that texture channel becomes deeper and texture details are clearer [16] with an increase in intensity factor of Improved G-L Fractional Differential mask. Improved G-L Fractional Differential mask can not only maintain the most energy of image on the low frequency, but also nonlinearly enhance its energy over intermediate and high frequency, which leads to richer texture details. For quantity analysis, information entropy and average gradient are taken as parameters, which are shown in Table II.

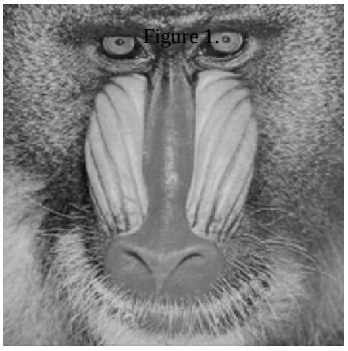
From Table II it is observed that with an increase in intensity factor of Fractional Differential mask, the information entropy and average gradient of the image start increasing. The information entropy reaches its maximum value at intensity factor $\gamma = 1.8$, thus nonlinearly preserve the low-frequency contour feature in the smooth area to the high degree. But as the intensity factor is increased beyond 1.8, the information entropy starts decreasing resulting in loss of textural information.

Intensity Factor	Information Entropy	Average Gradient	Intensity Factor	Information Entropy	Average Gradient
0	7.1677	9.0446	1.5	7.6823	27.6426
1.0	7.5716	21.2197	1.6	7.6923	28.9346
1.1	7.6035	22.4979	1.7	7.6975	30.2284
1.2	7.6288	23.7798	1.8	7.6991	31.5237
1.3	7.6518	25.0648	1.9	7.6963	32.8204
1.4	7.6695	26.3525	2.0	7.6885	34.1183

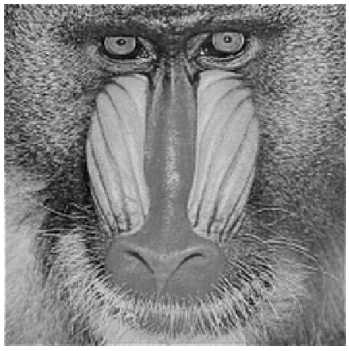
VI. Conclusion

In this paper, an improved G-L Fractional Differential mask is presented, which can enhance both texture and lightness of image. The Fractional Differential mask presented can control the degree of texture enhancement with Fractional Differential order v and intensity factor γ. And from quantitative analysis, it is observed that improvement of information entropy of image through improved G-L FD mask is more than both G-L and R-L FD Masks, thus enhancing more textural information. However, through all the FD Masks the image gets distorted if the value of intensity factor and fractional differential order is exceeded above certain value. The FD mask discussed, can be implemented further on real time images like remote sensed images, fingerprint images collected from the crime scene, etc., to enhance texture of these images, which is the subject of future work.

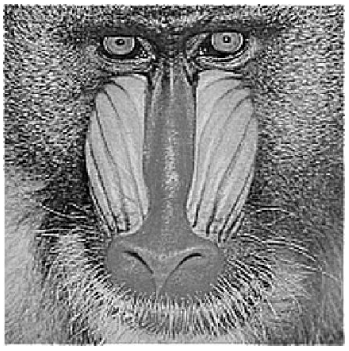
Figure 1.



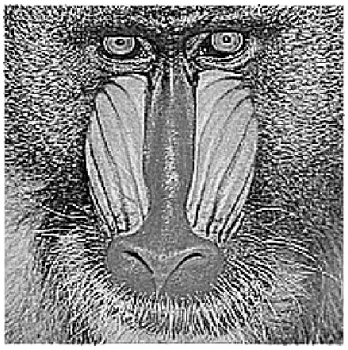
(a) v = 0



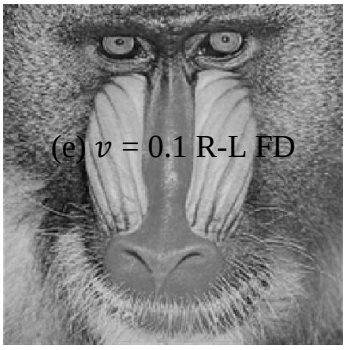
(b) v = 0.1 G-L FD



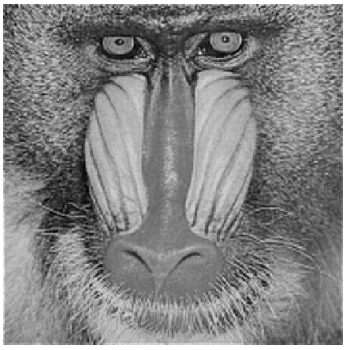
(c) v = 0.4 G-L FD



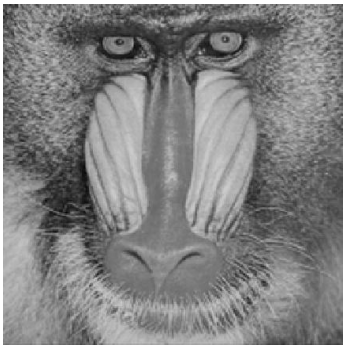
(d) v = 1 G-L FD



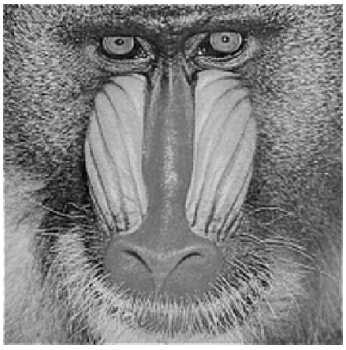
(f) v = 0.1 G-L FD



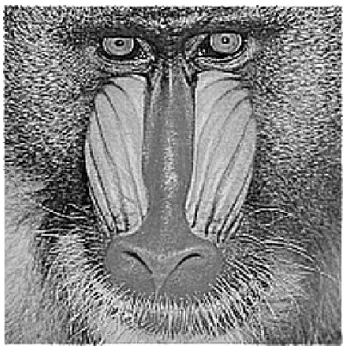
(g) v = 0.4 R-L FD



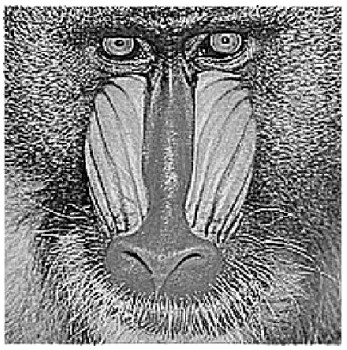
(h) v = 1 R-L FD



(i) v = 0.1 Improved G-L FD



(j) v = 0.4 Improved G-L FD

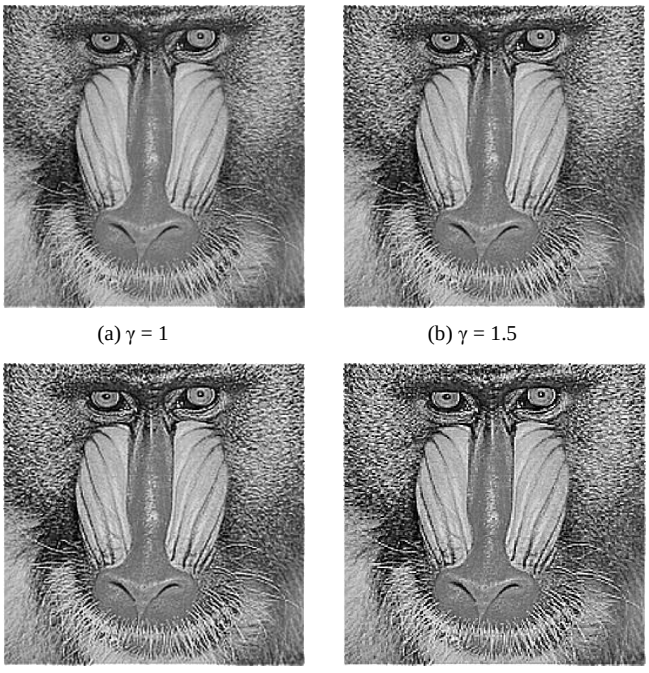


(k) v = 1 Improved G-L FD

Figure 2: Texture enhancement of Baboon Image with intensity factor γ = 1 for different values of fractional order (v) through different types of FD masks.

TABLE I. INFORMATION ENTROPY AND AVERAGE GRADIENT FOR VARIOUS IMAGES PROCESSED THROUGH DIFFERENT FD MASKS

Quantity / FD Order	Information Entropy									Average Gradient								
	G-L FD Mask			R-L FD Mask			Improved G-L FD Mask			G-L FD Mask			R-L FD Mask			Improved G-L FD Mask		
	Baboon	Bridge	Moon Surface	Baboon	Bridge	Moon Surface	Baboon	Bridge	Moon Surface	Baboon	Bridge	Moon Surface	Baboon	Bridge	Moon Surface	Baboon	Bridge	Moon Surface
0	7.1686	7.7285	7.0331	7.1686	7.7285	7.0331	7.1686	7.7284	7.0331	9.1536	8.3962	10.3749	9.1536	8.3962	10.3749	9.1536	8.3962	10.3749
0.1	7.3320	7.7657	7.1349	7.2285	7.7443	7.0562	7.3788	7.7744	7.1698	13.8149	10.3266	14.9990	11.1102	9.2084	12.3481	15.0309	10.8352	16.1845
0.2	7.5074	7.8007	7.2687	7.2837	7.7560	7.0965	7.5715	7.8136	7.3197	18.8088	12.4448	19.8604	12.5891	9.8176	13.8012	21.2196	13.4902	22.2057
0.3	7.6310	7.8248	7.3660	7.3200	7.7634	7.1258	7.6789	7.8335	7.4085	23.9133	14.6722	24.8279	13.5149	10.2014	14.7061	27.3150	16.1821	28.1421
0.4	7.6932	7.8351	7.4240	7.3350	7.7663	7.1378	7.6947	7.8324	7.4456	29.0690	16.9668	29.8521	13.9022	10.3630	15.0842	33.1318	18.7966	33.8149
0.5	7.6899	7.8297	7.4505	7.3309	7.7652	7.1341	7.6559	7.8146	7.4495	34.2533	19.3043	34.9093	13.8020	10.3212	14.9864	38.5195	21.2432	39.0734
0.6	7.6455	7.8104	7.4483	7.3110	7.7615	7.1182	7.5882	7.7877	7.4325	39.4552	21.6699	39.9870	13.2892	10.1077	14.4857	43.3330	23.4429	43.7740
0.7	7.5612	7.7783	7.4247	7.2785	7.7552	7.0918	7.5064	7.7558	7.4058	44.6685	24.0549	45.0785	12.4565	9.7627	13.6713	47.4248	25.3203	47.7712
0.8	7.4573	7.7335	7.3845	7.2399	7.7469	7.0618	7.4438	7.7260	7.3773	49.8898	26.4539	50.1797	11.4099	9.3316	12.6437	50.6421	26.8003	50.9149
0.9	7.3482	7.6807	7.3248	7.2008	7.7374	7.0411	7.3965	7.7035	7.3540	55.1167	28.8634	55.2880	10.2666	8.8612	11.5095	52.8257	27.8065	53.0489
1.0	7.2379	7.6206	7.2536	7.1686	7.7284	7.0331	7.3754	7.6940	7.3430	60.3477	31.2808	60.4017	9.1537	8.3962	10.3749	53.8095	28.2602	54.0104



(a) γ = 1 (b) γ = 1.5

(c) γ = 1.9 (d) γ = 2.0

Figure 3. Texture enhancement of Image with fractional order v = 0.2 for different values of intensity factor γ

ACKNOWLEDGMENT

The authors are thankful to unanimous reviewers for their valuable suggestions.

REFERENCES

[1] R. M. Haralick, K. Shanmugam and Its'Hak Dinstein, "Textural Features for Image Classification", IEEE Transaction on Systems, Man and Cybernetics, vol. 3, issue 6, pp. 610-621, 1973.

[2] Salem Saleh Al-amri, Dr. N. V. Kalyankar and Dr. S. D. Khamitkar, "Linear and Non-linear Contrast Enhancement Image", International Journal of Computer Science and Network Security, vol. 10, issue 2, pp. 139-143, 2010.

[3] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code", IEEE Transaction on Communication, vol. COM-31, issue 4, pp. 532–540, 1983.

[4] S. Dippel, M. Stahl, R. Wiemker, and T. Blaffert, "Multiscale contrast enhancement for radiographies: Laplacian pyramid versus fast wavelet transform", IEEE Transaction on Medical Imaging, vol. 21, issue 4, pp. 343–353, 2002.

[5] J. L. Starck, F. Murtagh, E. Candes, and D. Donoho, "Gray and color image contrast enhancement by the curvelet transform", IEEE Transaction on Image Processing, vol. 12, issue 6, pp. 706–717, 2003.

[6] M. N. Do and M. Vetterli, "The Contourlet transform: An efficient directional multiresolution image representation", IEEE Transaction on Image Processing, vol. 14, issue 12, pp. 2091–2106, 2005.

[7] E. R. Love, "Fractional derivatives of imaginary order", Journal of London Mathematical Society, vol. 3, pp. 241–259, 1971.

[8] K. B. Oldham and Spanier, The Fractional Calculus: Integrations and Differentiations of Arbitrary Order. New York: Academic, 1974.

[9] K. Nishimoto, Fractional Calculus. New Haven, CT: Univ. New Haven Press, 1989.

[10] Yi-Fei Pu, "Fractional Calculus Approach to Texture of Digital Image", 8th International Conference on Signal Processing, 2007.

[11] Zhuzhong Yang, Fangnian Lang, Xiaohong Yu and Yu Zhang, "The Construction of Fractional Differential Gradient Operator", Journal of Computational Information Systems, vol. 7, issue 12, pp. 4328-4342, 2011.

[12] Yawei Liu, "Remote Sensing Image Enhancement Based on Fractional Differential", International Conference on Computational and Information Science, pp. 881-884, 2010.

[13] Yi Zhang, Yi-Fei Pu and Jiliu Zhou, "Construction of Fractional differential Masks Based on Riemann-Liouville Definition", Journal of Computational Information Systems, vol. 6, issue 10, pp. 3191-3199, 2010.

[14] Zhifeng Gan and Hongyu Yang, "Texture Enhancement though Multiscale Mask based on RL Fractional Differential", International Conference on Information Networking and Automation, vol. 1, pp. 333-337, 2010.

[15] Yi-Fei Pu, Zhou Ji-Liu, and Yuan Xiao, "Fractional Differential Mask: A Fractional Differential-Based Approach for Multiscale Texture Enhancement", IEEE Transactions on Image Processing, vol. 19, issue 2, pp. 491-511, 2010.

[16] Gao Chaobang, Zhou Ji-Liu, Zheng Xiuqing and Lang Fangnian, "Image Enhancement Based on Improved Fractional Differentiation", Journal of Computational Information System, vol. 7, issue 1, pp. 257-264, 2011.

AUTHORS PROFILE

Vishwadeep Garg was born in Bathinda (Punjab). He received his B.Tech in Electronics and Communication Engineering from PTU, Jalandhar. He has completed his M.E in Electronics and Communication Engineering from Thapar University, Patiala. His research interest includes Digital Image Processing, and Digital Signal Processing.

Dr. Kulbir Singh was born in Batala (Pb) India. He received his BTech degree in 1997 from PTU, Jalandhar. He obtained his ME and Ph.D. degree from Thapar Institute of Engineering and Technology, Patiala in 2000 and 2006 respectively. He worked as lecturer from 2000 to 2007, Assistant Professor from 2007 to 2010 in Electronics and Communication Engineering Department, Thapar University, Patiala. Presently he is working as Associate Professor at Thapar University, Patiala since June 2010. He has published about 65 research articles in refereed international and national journals, international conference and national conference. He is life time member of IETE and ISTE. His research interests include Digital Signal Processing, Image Processing, Fractional Fourier Transform and Communication systems.

A Comparative Study on Temporal Mobile Access Pattern Mining Methods

Hanan Fahmy
Information System Dep,
Faculty of Computer and Information Systems, Helwan University,
Cairo, Egypt

Maha A.Hana
Information System Dep,
Faculty of Computer and Information Systems, Helwan University,
Cairo, Egypt

Yahia K. Helmy
Information System Dep,
Faculty of Computer and Information Systems, Helwan University,
Cairo, Egypt

***Abstract*—Mobile users behavior patterns is one of the most critical issues that need to be explored in mobile agent systems. Recently the algorithms of discovering frequent mobile user's behavior patterns have been studied extensively. Existing mining methods have proposed frequent mobile user's behavior patterns statistically based on requested services and location information. Therefore, other studies considered that the mobile user's dynamic behavior patterns are usually associated with temporal access patterns. In this paper, temporal mobile access pattern methods are studied and compared in terms of complexity and accuracy. The advantages and disadvantages of these methods will be summarized as well.**

Keywords- mobile mining; temporal data mining; mobile services; access pattern.

I. INTRODUCTION

The mobile industry has experienced significant growth during the past two decades. Practically the fast expansibility of ubiquitous mobile [1] technology has created a numberless opportunity to gather and extract information from mobile agent systems. Recently a critical problem was exist which is how services and access methods can be provided to mobile users. Mobile services are by definition consumed through a mobile handset, which is defined to mean a pocket-sized device with at least cellular connectivity capabilities. In the above mentioned definition of the mobile industry, network vendors, handset vendors, network operators, virtual network operators, service providers, content aggregators, third party software developers and emerging mobile Internet players are all therefore part of the mobile communications industry [2].

In the ubiquitous mobile computing environments, the mobile users may request diverse kinds of services and applications by mobile devices from arbitrary locations at any time via on networks [3]. Obviously, the behavior pattern, in which the location and the service are inherently coexistent with temporal associated, of mobile users becomes more complex than that of the traditional mobile agent systems. To assist the user get interested information on time is one of the promising applications, especially in mobile agent systems. Wherefore, effective modeling of mobile user behavior patterns is becoming very important. As effective modeling the behavior patterns of users in mobile agent systems benefits not only the users in smart access by caching or pre-fetching [4] [5] but also the mobile service providers to enhance their services and provide new services that may attract more subscribers [6].

There are existed studies that considered only one of the characteristics, i.e. location associated with requested services. Obviously, both movement and location or service requested with the temporal association rule should be considered simultaneously in order to discover complete information of user behavior patterns when the user request services. As a result, recently there are studies exist that discover the mobile user's interesting movement behavior patterns by temporal mobile access patterns in which this paper considered. These methods are Temporal Mobile access pattern (TMAP) [7] and Temporal mobile sequential pattern (TMSP) [8] as they have advantages and disadvantages of their own.

This paper aims to introduce a comparative study of the tow different methods of discovering temporal mobile behavior as these methods are studied respectively and compared in terms of complexity and accuracy. Furthermore, we explore the advantages and disadvantages of each of them.

The remainder of this paper is organized as follows. In Section 2, we gave a brief definition of mobile service. In Section 3, we briefly define the temporal data mining. Section 4 describes the comparative temporal mining methods in details with introducing the strength and weakness points of each one. Finally, the conclusion is given in Section 5.

II. MOBILE SERVICES

Mobile services differ from traditional services in their ability to provide service offerings regardless of temporal and spatial constraints. They are also different from traditional interpersonal services that deliver face-to-face, or from other types of e-services, such as wireless online services, where the service delivery linked to a specific fixed local area network or specific location. Mobile services have some special characteristics in comparison to other types of services [9].

The key differences related to spatial and temporal components of service usage [10]. For example, if one wants to meet a bank teller, she has to visit the bank location at a certain appointment time. These restrictions are present to some extent even with many electronic services. Even though online banking over a DSL Internet connection. Even though online banking services are available 24 hours a day thus overcoming

the problem of temporal availability, a fixed location is still needed for the DSL line, which is a spatial restriction. Mobile services, used with handheld mobile devices, overcome both spatial and temporal constraints [2] [9] an additional unique dimension of mobile services is the potential for individual personalization of service offerings.

Mobile handsets are multi-purpose private computers. This provides an attractive setting for electronic service delivery. Evaluation of mobile services as perceived by end-users is therefore challenging. Some papers claim that the opportunity cost of time is what matters [11]. Some analyses also recognize the dimension of level of effort in using the service [12].

III. TEMPORAL DATA MINING

Data mining is defined as the process of nontrivial extraction of previously unknown and useful knowledge from data. It discovers patterns hidden in data and associations between the patterns. Temporal data mining deals with the problem of mining patterns from temporal data, which can be either symbolic sequences or numerical time series [13]. Temporal data mining has the capability to look for interesting correlations or rules in large sets of temporal data.

The definition of temporal data mining presented in [14] is as the following. Temporal Data Mining is a single step in the process of Knowledge Discovery (KD) in temporal databases that enumerates structures (temporal patterns or models) over the temporal data, and any algorithm that enumerates temporal patterns from, or fits models to, temporal data is a temporal data mining algorithm.

Currently, temporal data mining is a fast expanding field with many research results reported. As a result, many new temporal data mining analysis methods or prototypes developed recently [15]. Two factors contribute to the popularity of temporal data mining. The first factor is an increase in the volume of temporal data stored, as many real-world applications deal with huge amount of temporal data. The second factor is the mounting recognition in the value of temporal data [16].

In many application domains, temporal data are now being viewed as invaluable assets from which hidden knowledge can be derived, to help understand the past and/or plan for the future [17].

IV. TEMPORAL MINING METHODS

In This section, we describe the temporal mobile access pattern mining methods, which are T-MAP mine and TMSP mine. These mining algorithms will be described in details respectively with their mining mechanisms followed by expressing the strength and weakness points of each mining method.

A. TMAP Mine

Temporal Mobile Access Patterns is a data mining method to discover mobile user's temporal behavior patterns associated with location and request services by using temporal association rule. Furthermore, it used data structure with temporal mobile access patterns called T-Map-Tree [7]. As shown in Fig 1 the workflow of TMAP mining mechanism contains two phases, which are data integration phase and mining phase. Data integration phase, is to collect and integrate users' logs into one dataset. Mining phase is to discover the frequently temporal mobile access patterns (T-Map) from the integrated log dataset [18].

TMAP use predefined timestamps by setting the time interval every four hours then they integrate the mobile user's current location information and service request into the integrated log as time goes by according to TMAP predefined time interval. The TMAP mine is consisted of a header table and aggregating the access patterns into the memory in a compact form. The head table are stored the frequently occurrence of mobile access patterns in the order of descending sequence access patterns. TMAP method needs one physical database scan when constructing of the header table. The database scan is to find all frequent mobile access patterns. Then, the frequent mobile access pattern is inserted into the header table in decreasing order of their sequence access pattern After the scan of the database, the set of frequent access patterns is sorted in the order of descending sequence data then the TMAP structure is constructed.

TMAP is conducted to efficiently find the mobile users' temporal mobile access pattern in distinct time interval, which is helpful to provide real-time customized personal service. However, TMAP method lacks of flexibilities because for each time interval segmentation the start time and end time should be determined and not every data is suitable for same cutting method. Finding the best segmentation in not easy in advance, the predication rate will be influenced by the segmenting point of time interval.

Although TMAP method is a fast method due to using the compact data structure form, but the log data sets consume much memory when the information data sets are storing into the memory. TMAP method works especially well for context-awareness data sets in mobile agent systems however, TMAP is not applied on real datasets, and its performance is not evaluated.

B. TMSP Mine

Temporal Mobile Sequential Patterns is a data mining method for discovering the Temporal Mobile Sequential Patterns (TMSPs) of mobile users in Location Based Services (LBS) environments. Furthermore, it uses location prediction strategies to predict the next movement of mobile users by utilizing the discovered TMSPs [8]. As shown in Fig 2 the workflow of TMSP mining mechanism contains three phases: segmentation of mobile transactions, discovery of TMSPs, and prediction user's behaviors. The segmentation of mobile transactions phase use the statistical method to decide a suitable count of time segmenting points and base on it to obtain the most suitable position of time segmenting points by genetic algorithm (GA). In the discovery of TMSPs phase, the mobile logs of users are analyzed by a data mining algorithm to obtain the TMSPs for each time interval. Furthermore, the TMSPs are used to predict the next location and service of mobile users. In prediction phase, the most suitable pattern is picked up by the historical transaction log, moving path, and current time interval to provide LBS like in advance preparing the service which user might request in advance, and

recommending related services, and user can query related service as well.

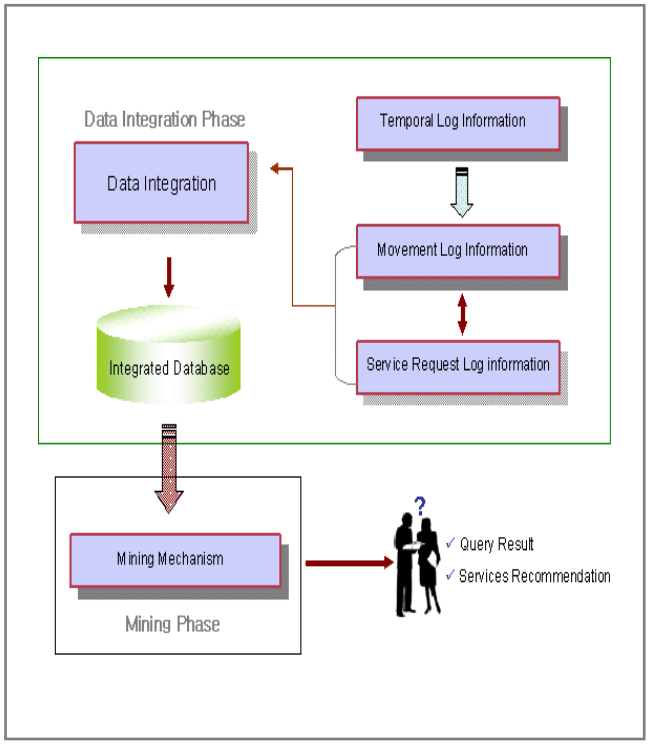


Figure 1. Work mechanisms for Temporal Mobile Access Pattern[18]

TMSP refer to TMAP-mine algorithm [8] without using time interval but using genetic algorithm to discover the most suitable time intervals by obtaining the most suitable positions of time segmenting point. TMSP-Mine algorithm determines the large transactions for each time interval and cell. TMSP-Mine algorithm uses the mapping table to transform the mobile transaction sequences into maximal large transaction sequences. By this step, TMSP gains advantage that can reduce the services of transactions that less than minimal support.

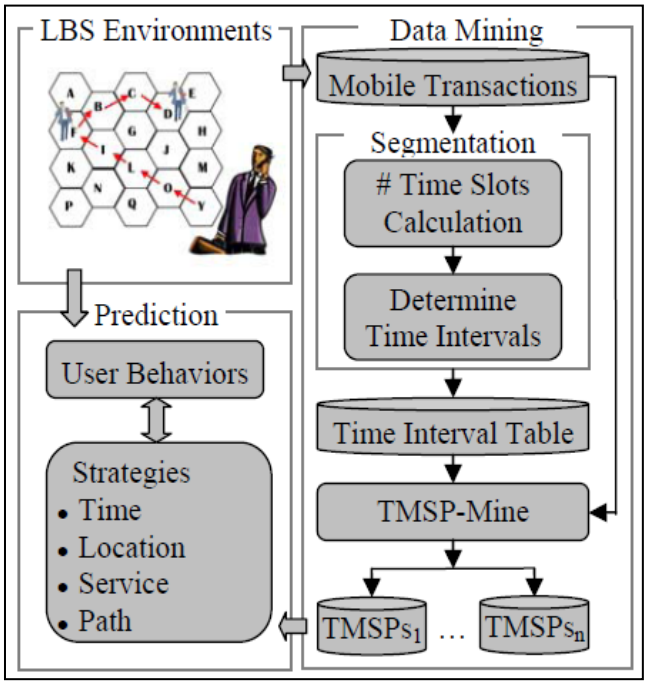
TMSP-Mine Algorithm utilizes TMSP-Tree, which is a two-level tree. The TMSP-Tree is well structured when comparing two patterns that have the same first and last transaction, and to generate candidate mobile sequential efficiently. This action is repeated more than one time until no candidate patterns can be generated which consume much time and memory space. TMSP use three prediction strategies for predicting behaviors of users by selecting the TMSPs based on the corresponding time interval; or with the longest length of fitting mobile user's previous location, service request, and moving path. If there are many TMSPs with the same length, TMSP is selected with the maximal support.

Although TMSP- mine method used for discovering the Temporal Mobile Sequential Patterns (TMSPs) of mobile users but the time segment is generated by searching into the time interval of the transaction log file that consume much time of searching; however it is no absolute assurance that a genetic algorithm will find a global optimum [19]. In addition, it store record for location without service in the dataset followed by record for location with service, which waste much memory space.

The genetic algorithm cannot assure constant optimization response times. Even more, the difference between the shortest and the longest optimization response time is much larger than with conventional gradient methods. This unfortunate genetic algorithm property limits the genetic algorithms' use in real time applications [20].

Figure 2. Work mechanisms for Temporal Mobile Sequential Pattern[8]

V. CONCLUSION



In this paper, we introduce a comparative study on temporal mobile behavior patterns after studying the existing mining methods. The two different temporal mobile access mining methods, which are TMAP and TMSP, are studied and compared in terms of complexity and accuracy. Our study shows that each one of them has their own advantages and disadvantages. Although the two algorithms used to predict the mobile behavior pattern in terms of service and location without concerning the dynamic time interval based of the response time of mobile service providers.

As a result, it remains an open question how to construct the best association rule based on sequential and temporal association rules for mobile users' access pattern for best recommend the next service that the mobile user will request in a dynamic timely manager according to the average response time of the mobile service providers. References

[1] Dan Lim,"Ubiquitous mobile Computing: UMC's model and success", Educational Technology & society 2(4) ISSN 1436-4522,1999.

[2] Hannu T. Verkasalo," Handset-Based Analysis of Mobile Service Usage",Phd Helsinki University of Technology, Faculty of Electronics, Communications and Automation, Department of Communications and Networking,2009.

[3] Yu ning, Hongbin Yang, " Sequence Mining for user behavior patterns in mobile Commerce", International Conference on Management of e-Commerce and e-Government, 978-0-7695-3366-7/08 $25.00, IEEE 2008.

[4] J.L.Chen," Resource allocation for cellular data services using multiagent schemes", IEEE Trans. Syst. Man Cybern. 31 (6) (2001) 864-869,2001.

[5] Vincent S. Tseng and W. C. Lin.,"Mining Sequential Mobile Access Patterns Efficiently in Mobile Web Systems", Proceeding of International Conference on Advanced Information Networking and Applications, pages 867-871,Taipei, Taiwan, March 2005.

[6] C.Y Chang, M.S Chen, "Integrating web caching and web prefetching in client-side proxies", in: Proceedings of the ACM 11th International conference Information and knowledge Management, 2002.

[7] S. C. Lee, J. Paik, J. Ok, I. Song, and U. M. Kim, "Efficient Mining of User Behaviors by Temporal Mobile Access Patterns". International Journal of Computer Science Security, Vol. 7, No. 2, pages 285-291, February 2007.

[8] Vincent S. Tseng, Eric Hsueh-Chan Lu, Cheng-Hsien Huang, "Mining Temporal Mobile Sequential Patterns in Location-Based Service Environments" , 978-1-4244-1890-9/07, IEEE 2007.

[9] Heinonen, K., Pura, M. , "Classifying Mobile Services", Proceedings of Helsinki Mobility Roundtable . Sprouts: Working Papers on Information Systems, 6(42). http://sprouts.aisnet.org/6-42, 2006.

[10] J. F. Roddick, K. Hornsby and M. Spiliopoulou, "Temporal Spatial and Spatio-Temporal data mining and knowledge discovery research Bibliography", http://kdm.first.flinders.edu.au/IDM/STDMBib.html.

[11] Goolsbee, A. & Klenow, PJ., "Valuing Consumer Products by the Time Spent Using Them: An Application to the Internet", Presented at the AEA Session on "The Roots of Innovation," Boston, Massachusetts, January 8, 2006.

[12] Pohjola, OP & Kilkki, K., "Value-based methodology to analyze communication services", Presented at Conference on Telecommunication Techno-Economics (CTTE) 2006.

[13] Dr. Naveeta Mehta et al. , " Temporal Sequential Pattern In Data Mining Tasks", International Journal on Computer Science and Engineering (IJCSE),2011.

[14] Lin, W., Orgun, M. A. & Williams, G. J. ,"An overview of temporal data mining", in S. J. Simoff, G. J. Williams & M. Hegland, eds, Proceedings of the 1st Australasian Data Mining Workshop (ADM'02)', Sydney, Australia, pp. 83–89,2002.

[15] Srikant R, Agrawal R: "Mining Sequential Patterns: Generalizations and Performance Improvements", in Int'l Conf Extending Database Technology. Springer 1996.

[16] Edi Winarko," The Discovery and Retrieval of Temporal Rules in Interval Sequence Data", Phd Flinders University, Faculty of Science and Engineering,2007.

[17] Chen, X. & Petrounias, I. ," A framework for temporal data mining", in 'Proceedings of the 9th International Conference on Database and Expert Systems Applications (DEXA'98)', Vienna, Austria, pp. 796–805, 1998.

[18] SeungCheol Lee, J. paik, J. Ok, I, Song and Ung Mo Kim, "Extracting Temporal Behavior Patterns of Mobile User", 978-0-7695-3322-3/08, IEEE 2008.

[19] William Sayers, "Genetic Algorithms and Neural Networks", Faculty of Advanced Technology, 05025397, milestone 3, Page 25-26,2009.

[20] Marczyk, A., "Genetic algorithms and evolutionary computatio"., http://www.talkorigins.com/faqs/genalg/genalg.htm, from Retrieved April13,2009.

A Schema for Generating Update Semantics

José Luis Carballido Carranza

Benemérita Universidad Autónoma de Puebla
jlcarballido7@gmail.com

Claudia Zepeda, Guillermo Flores

Benemérita Universidad Autónoma de Puebla
czepedac@gmail.com

***Abstract*—In this paper, we present a general schema for defining new update semantics. This schema takes as input any basic logic programming semantics, such as the stable semantics, the p-stable semantics or the MM^r semantics, and gives as output a new update semantics. The schema proposed is based on a concept called minimal generalized S models, where S is any of the logic programming semantics. Each update semantics is associated to an update operator. We also present some properties of these update operators.**

Keywords Update semantics, Logic Programming semantics, update properties.

I. Introduction

Updating, by definition means that there is new information that must be added to the older, and some information could be changed. Intelligent agents use this, in order to bring new knowledge to their knowledge base. But there is a main problem that updates can present, and it is inconsistency. So, it is important to use an approach to avoid inconsistencies in the knowledge base. For instance, it could be that in an initial moment we can infer a from a knowledge base (KB), and later the KB is updated with the new information $-a$ (where $-$ denotes negation). It is easy to see, that if we only take the union of the initial KB and $-a$, we will have an inconsistency. Then it is useful to apply an update approach that avoids the inconsistency and now allows to infer $-a$ since the newer knowledge has priority over the older. Currently there are several approaches in non-monotonic reasoning dealing with updates, such as [10, 16, 5].

As part of the contribution of this paper, we propose an schema for generating update semantics. This schema takes as input any basic logic programming semantics, such as the *stable semantics* [11], the *p-stable semantics* [17] or the MM^r semantics [13], and gives as output a new update semantics.

It is natural to consider the stable semantics since many approaches to updating have been based on it, see for example [10, 16, 5].

On the other hand, the p-stable semantics is another option to study updating. It has the advantage of providing models that coincide with classical models in many cases. We can make this clear with the following example. Let $P_1 = \{a \leftarrow \neg b, a \leftarrow b\}$ and $P_2 = \{b \leftarrow a\}$. From a classical logic point of view and considering that $\neg$ denotes classical negation, we would expect that $\{a, b\}$ corresponds to the result of updating P_1 with P_2. However, when we apply the approach in [10] based on stable semantics to update P_1 with P_2 there is no model; whereas our schema proposed with the p-stable semantics gives an update semantics that returns $\{a, b\}$ as an update model for the example.

Besides, in [14] it is shown that the p-stable semantics of normal programs can express any problem that can be expressed in terms of the stable semantics of disjunctive programs.

We also consider the MM^r semantics for several reasons. First, any normal program always has MM^r models. Second, it agrees with the Revised Stable models semantics defined by Pereira and Pinto for all the examples they present in their work [18], suggesting that both semantics may coincide for normal logic programs. The coincidence is important since the Revised Stable model semantics has the property of being a relevant semantics. One of the main implications of relevance is that it allows us to define top-down algorithms for answering queries from a knowledge base, this means that relevance allows us to split the original program into subprograms such that finding a model to answer a query can be reduced to finding the models of subprograms [18, 6, 7]. Third the MM^r has been used in the context of argumentation semantics, since it can identify the attack-dependencies that exist in an argumentation framework [13, 3].

Currently there exists a software implementation of the p-stable semantics and the MM^r semantics at `http://aplicacionesia.cs.buap.mx/~arkerz/` (Windows version) and at `http://sites.google.com/site/computingpstablesemantics/downloads` (Linux version).

The schema proposed is based on a concept called *minimal generalized S models*, where S is any of the mentioned logic programming semantics. The definition of minimal generalized S models is inspired by a concept called minimal generalized answer sets of abductive programs [12]. The semantics of minimal generalized answer sets is based on the stable semantics. The minimal generalized answer sets have been used to restore consistency [12, 2], to obtain the preferred plans of planning problems [21], to get the preferred extensions of an argument framework [21], and to define update operators [20]. Hence, we consider that minimal generalized S models can also have similar applications and be an alternative to those applications that use minimal generalized answer sets.

Each update semantics is associated to an update operator. Here we also present some properties of these update operators. These properties correspond to the properties of the update operator defined and analized by Eiter et al. [9] and J. J. Alferes et al. in [1], except for one of them called *independent parts property*. This last property refers to the general principle that asserts that completely independent parts of a program should not interfere with each other.

In section we summarize some basic concepts and definitions used to understand this paper. In section we review the minimal generalized S models. In section we present our schema for defining new update semantics and some formal properties. Finally, in section we present some conclusions.

II. Background

In this section, we define the syntax of the logic programs that we will use in this paper. In terms of logic programming semantics, we present the definition of the stable model semantics, the p-stable model semantics, and the MM^r semantics.

A. Logic programs

We use the language of propositional logic in the usual way. We consider *propositional symbols*: $p, q, \ldots$; *propositional connectives*: $\wedge, \vee, \rightarrow, \neg, -$; and *auxiliary symbols*: '(','),'.'. Well formed propositional formulas are defined as usual. We consider two types of negation: strong or classical negation (written as $-$) and negation-as-failure (written as $\neg$). Intuitively, $\neg a$ is true whenever there is no reason to believe a, whereas $-a$ requires a proof of the negated atom. An *atom* is a propositional symbol. A *literal* is either an atom a or the strong negation of an atom $-a$.

A *normal* clause is a clause of the form $a \leftarrow b_1 \wedge \ldots \wedge b_n \wedge \neg b_{n+1} \wedge \ldots \wedge \neg b_{n+m}$ where a and each of the b_i are atoms for $1 \leq i \leq n+m$. In a slight abuse of notation we will denote such a clause by the formula $a \leftarrow \mathcal{B}^+ \cup \neg\mathcal{B}^-$ where the set $\{b_1, \ldots, b_n\}$ will be denoted by $\mathcal{B}^+$, and the set $\{b_{n+1}, \ldots, b_{n+m}\}$ will be denoted by $\mathcal{B}^-$. Given a normal clause $a \leftarrow \mathcal{B}^+ \cup \neg\mathcal{B}^-$, denoted by r, we say that $a = H(r)$ is the *head* and $\mathcal{B}^+(r) \cup \neg\mathcal{B}^-(r)$ is the *body* of the clause.

A clause with an empty body is called a *fact*; and a clause with an empty head is called a *constraint*. Facts and constraints are also denoted as $a \leftarrow$ and $\leftarrow \mathcal{B}^+ \cup \neg\mathcal{B}^-$ respectively. We define a *normal logic program* P, as a finite set of normal clauses. The signature of a normal logic program P, denoted as $\mathcal{L}_P$, is the set of atoms that occur in P. Given a set of atoms M and a signature $\mathcal{L}$, we define $\neg\widetilde{M} = \{\neg a \mid a \in \mathcal{L} \setminus M\}$. Since we shall restrict our discussion to propositional programs, we take for granted that programs with predicate symbols are only an abbreviation of the ground program. From now on, by *program* we will mean a normal logic program when ambiguity does not arise.

In our programs we will manage the strong negation $-$ as follows: each atom $-a$ is replaced by a new atom symbol a' which does not appear in the language of the program and we add the constraint $\leftarrow a \wedge a'$ to the program.

B. Logic programming semantics

Here, we present the definitions of three logic programming semantics. Note that we only consider 2-valued logic programming semantics.

Definition 1. *A logic programming semantics S is a mapping from the class of all programs into the power set of the set of (2-valued) models.*

We sometimes refer to *logic programming semantics* as *semantics*, when no ambiguity arises. The semantics that we consider in this paper are: the MM^r semantics [13] that is based on the *the minimal model semantics* (denoted by MM), *the stable model semantics* [11] (denoted by *stable*), and *the p-stable model semantics* [17] (denoted by *p-stable*). We will review these semantics in the next subsections. From now on, we assume that the reader is familiar with the notion of an *interpretation* and *validity* [19].

When considering any particular semantics of a normal program with constraints $P \cup R$ (R is the set of constraints), we will understand the models given by that semantics of the program P that make the clauses of R valid in the sense of classical logic.

Stable semantics

The stable semantics was defined in terms of the so called *Gelfond-Lifschitz reduction* [11] and it is usually studied in the context of syntax dependent transformations on programs. The following definition of a stable model for normal programs was presented in [11].

Definition 2. *Let P be any program. For any set $M \subseteq \mathcal{L}_P$, let P^M be the definite program obtained from P by deleting each rule that has a literal $\neg l$ in its body with $l \in M$, and then all literals $\neg l$ in the bodies of the remaining clauses. Clearly P^M does not contain $\neg$, then M is a stable model of P if and only if M is a minimal model of P^M.*

Example 3. *Let $M = \{b\}$ and P be the following program: $\{b \leftarrow \neg a,\ \ c \leftarrow \neg b,\ \ b \leftarrow,\ \ c \leftarrow a\}$. Notice that P^M has three models: $\{b\}$, $\{b, c\}$ and $\{a, b, c\}$. Since the minimal model among these models is $\{b\}$, we can say that M is a stable model of P.*

p-stable semantics

Before defining the p-stable semantics (introduced in [17]), we define some basic concepts. Logical inference in classic logic is denoted by $\vdash$. Given a set of proposition symbols S and a theory (a set of well-formed formulas) Γ, $\Gamma \vdash S$ if and only if $\forall s \in S$, $\Gamma \vdash s$. When we treat a program as a theory, each negative literal $\neg a$ is regarded as the standard

negation operator in classical logic. Given a normal program P, if $M \subseteq \mathcal{L}_P$, we write $P \Vdash M$ when: $P \vdash M$ and M is a classical 2-valued model of P.

The p-stable semantics is defined in terms of a single reduction which is defined as follows:

Definition 4. *[17] Let P be a program and M be a set of literals. We define*

$$RED(P, M) = \{a \leftarrow \mathcal{B}^+ \cup \neg(\mathcal{B}^- \cap M) \mid a \leftarrow \mathcal{B}^+ \cup \neg\mathcal{B}^- \in P\}$$

Example 5. *Let us consider the program $P_1 = \{a \leftarrow \neg b \wedge \neg c,\ \ a \leftarrow b,\ \ b \leftarrow a\}$ and the set of atoms $M_1 = \{a, b\}$. We can see that $RED(P, M)$ is: $\{a \leftarrow \neg b,\ \ a \leftarrow b,\ \ b \leftarrow a\}$.*

Next we present the definition of the p-stable semantics for normal programs.

Definition 6. *[17] Let P be a program and M be a set of atoms. We say that M is a p-stable model of P if $RED(P, M) \Vdash M$. We use p-stable to denote the semantics operator of p-stable models.*

Example 7. *Let us consider again P_1 and M_1 of Example 5. Let us verify whether M_1 is a p-stable model of P_1. First, we can see that M_1 is a model of P_1, i.e., for each clause C of P_1, M_1 evaluates C to true. We also can verify that $RED(P_1, M_1) \vdash M_1$. Then we can conclude that $RED(P_1, M_1) \Vdash M_1$. Hence, M_1 is a p-stable model of P_1.*

The following examples illustrate how to obtain the p-stable models. The first example shows a program with a single p-stable model, which is also a classical model. The second example shows a program which has no stable models and whose p-stable and classical models are the same.

Example 8. *Let $P = \{q \leftarrow \neg q\}$. Let us take $M = \{q\}$ then $RED(P, M) = \{q \leftarrow \neg q\}$. It is clear that M models P in classical logic and $RED(P, M) \models M$ since $(\neg q \rightarrow q) \rightarrow q$ is a theorem in classical logic with the negation $\neg$, now interpreted as classical negation. Therefore M is a p-stable model of P.*

Example 9. *Let $P = \{a \leftarrow \neg b,\ \ a \leftarrow b,\ \ b \leftarrow a\}$. We can verify that $M = \{a, b\}$ models the clauses of P in classical logic. We find that $RED(P, M) = P$. Now, from the first and third clause, it follows that $(\neg b \rightarrow b)$ where the negation $\neg$ is now interpreted as classical negation. Since $(\neg b \rightarrow b) \rightarrow b$ is a theorem in classical logic, it follows that $RED(P, M) \models M$. Therefore, M is a p-stable model of P.*

It is worth mentioning that there exists also a characterization of the p-stable semantics in terms of the paraconsistent logic G'_3, interested readers can see [14, 15, 17].

Minimal model semantics

An interpretation M is called a (2-valued) *model* of P if and only if for each clause $c \in P$, $M(c) = 1$. We say that M is a *minimal model* of P if and only if there does not exist a model M' of P such that $M' \subset M$, $M' \neq M$ [19]. We will denote by $MM(P)$ the set of all the minimal models of a given logic program P. Usually MM is called *minimal model semantics*.

Example 10. *Let P be the program $\{a \leftarrow \neg b,\ \ b \leftarrow \neg a,\ \ a \leftarrow \neg c,\ \ c \leftarrow \neg a\}$. As we can see, P has five models: $\{a\}$, $\{b, c\}$, $\{a, c\}$, $\{a, b\}$, $\{a, b, c\}$; however, P has just two minimal models: $\{b, c\}$, $\{a\}$. Hence $MM(P) = \{\ \{b, c\}, \{a\}\ \}$.*

The MM^r semantics

A program P induces a notion of *dependency* between atoms from $\mathcal{L}_P$ [13]. We say that *a depends immediately on b*, if and only if, b appears in the body of a clause in P, such that a appears in its head. The two place relation *depends on* is the transitive closure of *depends immediately on* [13]. The set of dependencies of an atom x, denoted by *dependencies-of*(x), corresponds to the set $\{a \mid x \text{ depends on } a\}$.

Example 11. *[13] Let us consider the following program,*
$P = \{e \leftarrow e,\ \ c \leftarrow c,\ \ a \leftarrow \neg b \wedge c,\ \ b \leftarrow \neg a \wedge \neg e,\ \ d \leftarrow b\}$.
The dependency relations between the atoms of $\mathcal{L}_P$ are as follows: dependencies-of$(a) = \{a, b, c, e\}$; dependencies-of$(b) = \{a, b, c, e\}$; dependencies-of$(c) = \{c\}$; dependencies-of$(d) = \{a, b, c, e\}$; and dependencies-of$(e) = \{e\}$.

We take $<_P$ to denote the strict partial order defined as follows: $x <_P y$, if and only if, *y depends-on x* and x does not *depend-on y*. By considering the relation $<_P$, each atom of $\mathcal{L}_P$ is assigned an order as follows: An atom x *is of order* 0, if x is minimal in $<_P$. An atom x *is of order* $n + 1$, if n is the maximal order of the atoms on which x depends.

We say that a program P is of order n, if n is the maximum order of its atoms. We can also break a program P of order n into the disjointed union of programs P_i with $0 \leq i \leq n$, such that P_i is the set of clauses for which the head is of order i (*w.r.t.* P). The empty program has order 0. We say that $P_0, \ldots, P_n$ are the *components* of P.

Example 12. *By considering the program P in Example 11, we can see that: d is of order 2, a is of order 1, b is of order 1, e is of order 0, and c is of order 0. This means that P is a program of order 2. The following table illustrates how the program P can be broken into the disjointed union of the following relevant modules or components $P_0 = \{e \leftarrow e,\ \ c \leftarrow c\}$, $P_1 = \{a \leftarrow \neg b \wedge c,\ \ b \leftarrow \neg a \wedge \neg e\}$, $P_2 = \{d \leftarrow b\}$.*

Next we present a reduction that will be used to define the MM^r semantics.

Let P be a program and $A = \langle T; F\rangle$ be a pair of disjoint sets of atoms. The reduction $R(P, A)$ is obtained by 2 steps [13]:

1. Let $R'(P, A)$ be the program obtained in the following steps:
   (a) We replace every atom x that occurs in the bodies of P by 1 if $x \in T$, and we replace every atom x that occurs in the bodies of P by 0 if $x \in F$;
   (b) we replace every occurrence of $\neg 1$ by 0 and $\neg\ 0$ by 1;
   (c) every clause with a 0 in its body is removed;
   (d) finally we remove every occurrence of 1 in the body of the clauses.

2. Given the set of transformations $\mathcal{CS} = \{RED^+, RED^-, Success, Failure, Loop\}$, in [8] it is shown that a normal program P can be reduced to another normal program $norm_{CS}(P)$ after applying those transformations a finite number of times. The program $norm_{CS}(P)$ is unique and is called the normal form of program P with respect to the system CS. We will denote $R(P, A) = norm_{CS}(R'(P, A))$. The definitions of the transformations in $\mathcal{CS}$ are:
   (a) If $r \in P$ and $a \in B^-(r)$ $\nexists r' \in P : H(r') = a$, then
   $$\mathbf{RED}^+(P) = (P \setminus \{r\}) \cup \{H(r) \leftarrow B^+(r) \cup \neg(B^-(r) \setminus \{a\})\}$$
   (b) If $r \in P$ and $a \leftarrow \in P$ such that $a \in B^-(r)$, then
   $$\mathbf{RED}^-(P) = P \setminus \{r\}$$
   (c) If $r \in P$ and $a \leftarrow \in P$ such that $a \in B^+(r)$, then
   $$\mathbf{Succes}(P) = (P \setminus \{r\}) \cup \{H(r) \leftarrow (B^+(r) \setminus \{a\}) \cup \neg B^-(r)\}$$
   (d) If $r \in P$ and $a \in B^+(r)$ $\nexists r' \in P : H(r') = a$, then
   $$\mathbf{Failure}(P) = P \setminus \{r\}$$
   (e) Let M be unique minimal model of the positive program
   $$\mathbf{POS}(P) = \{H(r) \leftarrow B^+(r) : r \in P\}$$
   then
   $$\mathbf{LOOP}(P) = \{r : r \in P, B^+(r) \subseteq M\}$$

Example 13. *[13] Let $Q = \{a \leftarrow \neg b \wedge c,\ b \leftarrow \neg a \wedge \neg e,\ d \leftarrow b,\ b \leftarrow e, m \leftarrow n, n \leftarrow m\}$, and let A be the pair of sets of atoms $\langle\{c\}; \{e\}\rangle$. Thus, $R'(Q, A) = \{a \leftarrow \neg b,\ b \leftarrow \neg a,\ d \leftarrow b, m \leftarrow n, n \leftarrow m\}$. Hence, $R(Q, A) = \{a \leftarrow \neg b,\ b \leftarrow \neg a,\ d \leftarrow b\}$.*

Now, in order to define the MM^r semantics, we first define the MM^r_c semantics in terms of the Minimal Model semantics, denoted by MM.

Definition 14. *Given $\boldsymbol{A} = \{A_1 \ldots A_n\}$ where the A_i, $1 \leq i \leq n$ are sets, and $\boldsymbol{B} = \{B_1 \ldots B_m\}$ where the B_j, $1 \leq j \leq m$ are sets, we define $\boldsymbol{A} \uplus \boldsymbol{B} = \{A_i \cup B_j \mid A_i \in \boldsymbol{A}$ and $B_j \in \boldsymbol{B}\}$.*

Definition 15. *[13] We define the associated MM^r_c semantics recursively as follows: Given a program P of order 0, $MM^r_c(P) = MM(P)$. For a program P of order $n > 0$ we define*

$$MM^r_c(P) = \bigcup_{M \in MM(P_0)} \{M\} \uplus MM^r_c(R(P \setminus P_0, \langle M; N\rangle))$$

where $N := (\mathcal{L}_{P_0} \cup \{a \in \mathcal{L}_P \mid a \notin Head(P)\}) \setminus M$.

It is important to eliminate tautologies from the programs, since they can introduce non-desirable models. For example, if P is the program $\{a \leftarrow \neg b,\ b \leftarrow a, b\}$, then the minimal models for this program are $\{a\}$ and $\{b\}$; however, after deleting the second rule, which is a tautology, it is clear that the second set, namely $\{b\}$ is not an intended minimal model. Given a normal program P, we define $Taut(P) = \{a \leftarrow \mathcal{B}^+ \cup \neg\mathcal{B}^- \in P \mid \mathcal{B}^+ \cap \mathcal{B}^- \neq \emptyset$ *or* $a \in \mathcal{B}^+\}$.

Definition 16. *[13] Let P be a normal program. We define $MM^r(P) = MM^r_c(P \setminus Taut(P))$.*

The following example illustrates our two previous definitions.

Example 17. *Let us consider the program $E = \{a \leftarrow \neg b,\ b \leftarrow \neg a,\ p \leftarrow \neg b,\ p \leftarrow \neg p\}$. We are going to compute the $MM^r_c(E)$. According to Definition 15, since E is of order 1, then we need to obtain the following:*
1) $MM(E_0)$,
2) $MM^r_c(R(E \setminus E_0, \langle M; N\rangle))$ for each $M \in MM(E_0)$, and
3) $MM^r_c(E) = \bigcup_{M \in MM(E_0)} \{M\} \uplus MM^r_c(R(E \setminus E_0, \langle M; N\rangle))$.

***Obtaining** $MM(E_0)$: Let us see that $E_0 = \{a \leftarrow \neg b,\ b \leftarrow \neg a\}$ is the component of order 0 of program E. Thus $MM(E_0) = MM(E_0) = \{\{a\}, \{b\}\}$.*

***Obtaining** $MM^r_c(R(E \setminus E_0, \langle M; N\rangle))$ **for each** $M \in MM(E_0)$: There are two cases to consider.*

Let us consider M to be $\{a\}$. Then E' is the program $R(E \setminus E_0, \langle M; N\rangle)$ with $E \setminus E_0 = \{p \leftarrow \neg b,\ \ p \leftarrow \neg p\}$, and $N = \{b\}$. We can see that $E' = \{p \leftarrow,\ \ p \leftarrow \neg p\}$. Now we need to obtain $MM^r_c(E')$ which is the same as $MM(E') = \{\{p\}\}$. Hence, $\{M\} \uplus MM^r_c(E') = \{\{a\}\} \uplus \{\{p\}\} = \{\{a, p\}\}$.

Let us consider M to be $\{b\}$. Let E' be the program $R(E \setminus E_0, \langle M; N\rangle)$ with $E \setminus E_0 = \{p \leftarrow \neg b,\ \ p \leftarrow \neg p\}$, and $N = \{a\}$. We can see that $E' = \{p \leftarrow \neg p\}$. Now we need to obtain $MM^r_c(E')$ which is the same as $MM(E') = \{\{p\}\}$. Hence, $\{M\} \uplus MM^r_c(E') = \{\{b\}\} \uplus \{\{p\}\} = \{\{b, p\}\}$.

***Obtaining** $MM^r_c(E)$: It is easy to verify that $MM^r_c(E) = \bigcup_{M \in MM(E_0)} \{M\} \uplus MM^r_c(R(E \setminus E_0, \langle M; N\rangle)) = \{\{a, p\}, \{b, p\}\}$.*

Since E is a program with no tautologies then $MM^r(E) = MM^r_c(E)$.

Some properties of logic programming semantics

Not all normal programs have p-stable models or stable models [17, 11], although they always have MM^r models [13], that is why it is convenient to have the next definition.

Definition 18. *Let S be any of the three semantics. Let P be a program. We say that P is S consistent if P has at least one S model. We say that P is S inconsistent if P does not have S models.*

Now, we present two notions of equivalence for programs.

Definition 19. *[4] Let S be any of the three semantics. Two programs P_1 and P_2 are equivalent, denoted by $P_1 \equiv_S P_2$, if P_1 and P_2 have the same S models. Two programs P_1 and P_2 are strongly equivalent, denoted by $P_1 \equiv_{SE(S)} P_2$, if $(P_1 \cup P) \equiv_S (P_2 \cup P)$ for every program P. We will drop the subindex S that follows the equivalent symbol whenever no ambiguity arises.*

The following lemma[1] indicates that given a program P and an atom x that does not occur in P, we can define a new program P' such that P and P' are equivalent and $\mathcal{L}_{P'} = \mathcal{L}_P \cup \{x\}$. The two programs must have the same clauses except for one of them. One of the clauses in P' corresponds to one of the clauses in P after adding $\neg x$ to its body. This way, P and P' have the same S models since x does not appear as the head of any clause in P'.

Lemma 20. *Let S be any of the three semantics. Let P be a program and x be an atom, $x \notin \mathcal{L}_P$. Let r be any clause $a \leftarrow \mathcal{B}^+ \cup \neg\mathcal{B}^-$ in P. Then M is a S model of P iff M is a S model of $(P \setminus \{r\}) \cup \{a \leftarrow \mathcal{B}^+ \cup \neg(\mathcal{B}^- \cup \{x\})\}$.*

III. Minimal generalized S models

The definition of our schema for generate update semantics is based on a concept called Minimal Generalized S models, denoted as *MG S models*, where S is any of the three semantics given in the Section , namely *stable*, $p-stable$, or MM^r semantics.

The intuition behind the MG S models is simple. Given a semantics S, a program P and a set of atoms A, the MG S models of P are the S models of $P \cup \Delta$ that are obtained by adding the minimal subset $\Delta \subseteq A$ to P for which $P \cup \Delta$ has S models.[2] For instance, let us consider the program $P = \{-a,\ a \leftarrow \neg b\}$, $A = \{b, c\}$, and the p-stable semantics, hence $\{b, -a\}$ is one of its MG p-stable models where the minimal subset of A added to P is $\{b\}$. We also can see that that P does not have p-stable models.

Next, we present the definition of abductive logic programs and their semantics in terms of the *minimal explicit generalized S models*. Then, we define the MG S models based on the minimal explicit generalized S models. These definitions are similar to the definitions of syntax and semantics of abductive logic programs as presented in the context of the stable semantics in [2].

Definition 21. *Let S be a semantics. An abductive logic program is a pair $\langle P, A\rangle$ where P is a program and A is a set of atoms, called abducibles. $\langle M, \Delta\rangle$ is an explicit generalized S model, denoted as EG S model, of the abductive logic program $\langle P, A\rangle$ iff $\Delta \subseteq A$ and M is an S model of $P \cup \Delta$.*

We give an ordering among EG S models in order to get the minimal of them.

Definition 22. *Let S be a semantics. Let $T = \langle P, A\rangle$ be an abductive logic program. Let $\langle M_1, \Delta_1\rangle$ and $\langle M_2, \Delta_2\rangle$ be two EG S models of T, we define $\langle M_1, \Delta_1\rangle < \langle M_2, \Delta_2\rangle$ if $\Delta_1 \subset \Delta_2$; this order is called inclusion order. $\langle M, \Delta\rangle$ is a Minimal EG S model, denoted as MEG S model, of T iff $\langle M, \Delta\rangle$ is an EG S model of T and it is minimal w.r.t. inclusion order.*

For practical purposes, given a MEG S model, $\langle M, \Delta\rangle$, we are only interested in its first entry, namely M, and we call it a *Minimal Generalized S model*, denoted as *MG S model*, of an abductive logic program.

Example 23. *Let S be the p-stable semantics. Let $\langle P, A\rangle$ be the abductive logic program where the set of abductive atoms is $A = \{x_1,\ x_2\}$ and $P = \{b \leftarrow \neg x_1,\ a \leftarrow b \wedge \neg x_2,\ -a\}$. There are three EG p-stable models of $\langle P, A\rangle$ which are: $\langle\{-a, x_1\}, \{x_1\}\rangle$, $\langle\{-a, b, x_2\}, \{x_2\}\rangle$, and $\langle\{-a, x_1, x_2\}, \{x_1, x_2\}\rangle$. We can see that for $\Delta = \emptyset$ there is no EG p-stable models. Therefore, the MEG p-stable models are $\langle\{-a, x_1\}, \{x_1\}\rangle$ and $\langle\{-a, b, x_2\}, \{x_2\}\rangle$, and the MG p-stable models are $\{-a, x_1\}$ and $\{-a, b, x_2\}$.*

The following lemma presents some results about MEG S models that will be useful in a later section to show the properties of our update operator. The proof of this lemma is straightforward.

Lemma 24. *Let $T = \langle P, A\rangle$ be an abductive logic program such that P is S consistent. Then,*

- *M is a S model of P iff M is a MG S-model of T and*
- *if $\langle M, \Delta\rangle$ is a MEG S model of T then $\Delta = \emptyset$.*

IV. Updates semantics and formal properties

In this section, we define the general schema for generate update semantics based on the concept of MG S models, and we study some of its properties. We use $\odot_S$ to represent the update operator with respect to a semantics S. In order to obtain the $\odot_S$-update models of a pair of logic programs (P_1, P_2), called *update pair*, we define an update logic program, denoted as P. The update logic program is obtained

[1] Its proof is straightforward.

[2] By "adding the minimal subset $\Delta \subseteq A$ to P", we mean that Δ is interpreted as a set of facts defined by its elements.

by joining P_1' to P_2, where P_1' is the resulting program from transforming P_1 as follows: at the end of each clause of P_1 which is not a constraint we add the negation-as-failure of an abducible (a new atom). The intuition behind the transformation applied to a program P_1 consists in weakening the knowledge in P_1 when giving more relevance to the knowledge contained in P_2 whose clauses are not modified.

Definition 25. *Let (P_1, P_2) be an update pair over $\mathcal{L}_{P_1 \cup P_2}$ such that the number of clauses in P_1 that are not constraints is n. Let $\mathcal{L}^*_{P_1 \cup P_2} = \mathcal{L}_{P_1 \cup P_2} \cup A$ where A is a set of n new abducible atoms, namely $A = \{a_i, 1 \leq i \leq n \mid a_i$ is an atom, $a_i \notin \mathcal{L}_{P_1 \cup P_2}$ and $a_i \neq a_j$ if $i \neq j\}$. We define the update logic program P of (P_1, P_2) over $\mathcal{L}^*_{P_1 \cup P_2}$, as the program consisting of the following clauses:*

1. *all constraints in P_1,*
2. *the clauses $a \leftarrow \mathcal{B}^+ \cup \neg(\mathcal{B}^- \cup \{a_i\})$ if $r_i = a \leftarrow \mathcal{B}^+ \cup \neg\mathcal{B}^- \in P_1$, $1 \leq i \leq n$ and $a_i \in A$,*
3. *all clauses $r \in P_2$.*

We define the abductive logic program of P as follows: $T = \langle \mathsf{P}, A \rangle$.

In this way, given a semantics S, the intended $\odot_S$-update models of a pair of logic programs (P_1, P_2) are obtained by removing the abducible atoms from the MG S models of the abductive logic program $\langle \mathsf{P}, A \rangle$. Finally, the $\odot_S$-update models are chosen as those that contain more information, i.e. maximal in the sense of inclusion of sets, from the intended $\odot_S$-update models.

Definition 26. *Let S be a semantics. Let (P_1, P_2) be an update pair over $\mathcal{L}_{P_1 \cup P_2}$ and T its abductive logic program. Then, $M \subseteq \mathcal{L}_{P_1 \cup P_2}$ is an intended $\odot_S$-update model of (P_1, P_2) if and only if $M = M' \cap \mathcal{L}_{P_1 \cup P_2}$ for some MG S model M' of T. In case M is an intended $\odot_S$-update model of (P_1, P_2) and is maximal among all intended $\odot_S$-update models of (P_1, P_2) w.r.t. inclusion order, then M is an $\odot_S$-update model of (P_1, P_2).*

We can illustrate our semantics with the following example.

Example 27. *Let S be the p-stable semantics. Let (P_1, P_2) be an update pair over $\{a, b\}$ where, P_1 and P_2 are the following logic programs, $P_1 = \{b \leftarrow, \quad a \leftarrow b\}$ and $P_2 = \{-a \leftarrow\}$. We can see that the update logic program P of (P_1, P_2) over $\mathcal{L}^*_{P_1 \cup P_2}$ corresponds to the program P of Example 23 where the x_i are the abducible a_i. The intended $\odot_{p-stable}$-update models of (P_1, P_2) are $\{-a\}$ and $\{-a, b\}$; and its only $\odot_{p-stable}$-update model is $\{-a, b\}$.*

Now, we show that our update operator $(\odot_S)$ satisfies several formal properties. These properties have been deeply analyzed, in the context of stable semantics, by several authors such as J. J. Alferes et al. in [1] or T. Eiter in [10], except for the last one. We will see that all the properties are expressed in terms of equivalence, hence it is useful to recall the two notions of equivalence for logic programs given in Definition 19. Since the S models of a logic program are sets of literals, we can see easily that $\equiv$ represents an equivalence relation, and the logic programs P_1 and P_2 can be of any kind defined in this paper.

The following definition is used to define the last of our properties.

Definition 28. *Let S be a semantics. Let (P_1, P_2) be a pair of logic programs over $\mathcal{L}_{P_1 \cup P_2}$. We define the update semantic function of (P_1, P_2) as follows:*

$$SEM_{\odot_S}(P_1, P_2)^3 = \{M \mid M \text{ is an } \odot_S \text{-update model of } (P_1, P_2)\}.$$

Now we define the properties for $\odot_S$ when S is the stable or p-stable semantics. In the case of the MM^r semantics these properties have not been verified, the study of them are the topic of future work. Since the intuition behind the first six properties is easy, we only give a deeper explanation about the last property below. For any of the semantics S we have the following properties.

P1. *Initialisation:* If P is a logic program then $\emptyset \odot_S P \equiv P$.

P2. *Strong consistency:* Let P_1 and P_2 be logic programs. Suppose $P_1 \cup P_2$ has at least one p-stable model. Then $P_1 \odot_S P_2 \equiv P_1 \cup P_2$.

P3. *Idempotence:* If P is a logic program then $P \odot_S P \equiv P$.

P4. *Weak noninterference:* If P_1 and P_2 are logic programs defined over disjoint alphabets, and both of them have p-stable models or do not, then $P_1 \odot_S P_2 \equiv P_2 \odot_S P_1$.

P5. *Weak irrelevance of syntax:* Let P, P_1 and P_2 be logic programs under $\mathcal{L}_{\mathsf{P}}$. If $P_1 \equiv_{SE} P_2$ then $P \odot_S P_1 \equiv P \odot_S P_2$.

P6. *Augmented update:* Let P_1 and P_2 be logic programs such that $P_1 \subseteq P_2$. Then $P_1 \odot_S P_2 \equiv P_2$.

P7. *Independent parts property.* Let $J_1 = (P_1, P_1')$ and $J_2 = (P_2, P_2')$ such that $(\mathcal{L}_{J_1} \cap \mathcal{L}_{J_2}) = \emptyset$. Then $SEM_{\odot_S}((P_1 \cup P_2), (P_1' \cup P_2')) = SEM_{\odot_S}(J_1) \uplus SEM_{\odot_S}(J_2)$.

Property **P7** indicates that our update operator does not violates the general principle that completely independent parts of a logic program should not interfere with each other. Hence the property **P7** of operator $\odot$ indicates that if we update the union of a pair of logic programs $(P_1 \cup P_2)$ by the union of a different pair of logic programs $(P_1' \cup P_2')$ such that P_1 and P_1' are defined under a different language from the language of logic programs P_2 and P_2' then, the

[3] Let us notice that $SEM_{\odot_S}(P_1, P_2)$ is a set of sets.

result can be also obtained from a particular union of the update of P_1 by P_1' and the update of P_2 by P_2'. This particular union of updates corresponds to our Definition 28.

The next example is taken from [16], where it is used for different purposes.

Example 29.

Let P_1 be:
$openSchool \leftarrow .$
$holiday \leftarrow \neg workday.$

Let P_1' be:
$-openSchool \leftarrow holiday.$
$workday \leftarrow \neg holiday.$

Let P_2 be:
$seeStars \leftarrow .$

Let P_2' be:
$-seeStars \leftarrow .$

Let S be the p-stable semantics. Let $J_1 = (P_1, P_1')$, $J_2 = (P_2, P_2')$, and $J = ((P_1 \cup P_2), (P_1' \cup P_2'))$. We can see that $(\mathcal{L}_{J_1} \cap \mathcal{L}_{J_2}) = \emptyset$.
According to independent parts property we have that,
$SEM_{\odot_{p-stable}}((P_1 \cup P_2), (P_1' \cup P_2')) = SEM_{\odot_{p-stable}}(J_1) \uplus SEM_{\odot_{p-stable}}(J_2)$ *since*
$SEM_{\odot_{p-stable}}((P_1 \cup P_2), (P_1' \cup P_2')) =$
$\{\{openSchool, workday, -seeStars\}\} =$
$\{\{openSchool, workday\}\} \uplus \{\{-seeStars\}\} =$
$SEM_{\odot_{p-stable}}(J_1) \uplus SEM_{\odot_{p-stable}}(J_2)$.

Theorem 30. *The update operator ($\odot_S$) satisfies properties,* **P1**, **P2**, **P3**, **P4**, **P5**, **P6**, *and* **P7** *when S is stable or p-stable semantics.*

Proof. We present the proof for the p-stable semantics. Properties **P1** to **P6** for stable semantics are proved in [20]. The proof of property **P7** for the stable semantics is similar to the one presented here for the p-stable semantics.

First, it is straightforward to verify that given a p-stable consistent program P, if M is p-stable model of P then there is not another p-stable model M' of P such that $M' \subset M$. So, by this last fact and by Lemma 24, it is also straightforward to verify that given an abductive logic program $\langle P, A\rangle$, where P is p-stable consistent, then if M is a MG p-stable model of $\langle P, A\rangle$ then there is not another MG p-stable model M' of $\langle P, A\rangle$ such that $M' \subset M$.
(**P1. Initialisation**): $\emptyset \odot P = P$ by construction. Hence $\emptyset \odot_{p-stable} P \equiv P$.
(**P2. Strong consistency**): Let $Q = (P_1 \cup P_2)$ such that Q is p-stable consistent. Let $J = (P_1, P_2)$. We must prove that M is an $\odot_{p-stable}$ update model of J iff M is a p-stable model of Q.
Let us notice that programs Q and P have the same clauses except for some of them, namely in P there are some clauses that have an abducible atom (a new atom) in their body and these atoms do not occur in Q. So when we apply iteratively Lemma 20, two things are certain:

(1) S is a p-stable model of Q, iff S is also a p-stable model of P, and

(2) if Q is p-stable consistent, then P is p-stable consistent too.

($\Rightarrow$) By hypothesis M is an $\odot_{p-stable}$ update model of J, then by Definition 26, there exists a MG p-stable model M' of the abductive logic program of J, $\langle \mathsf{P}, A\rangle$, such that $M = M' \cap \mathcal{L}_J$. Then by Definition 22, there exists Δ, $\Delta \subseteq A$ such that $\langle M', \Delta\rangle$ is a MG p-stable model of $\langle \mathsf{P}, A\rangle$, where $M = M' \cap \mathcal{L}_J$.

By hypothesis, Q is p-stable consistent then, by (2) P is p-stable consistent too. Hence applying Lemma 24, it is possible to verify that $\Delta = \emptyset$ and $M' = M$. So $\langle M, \emptyset\rangle$ is a MEG p-stable model of $\langle \mathsf{P}, A\rangle$. Finally by Definition 21, M is a p-stable model of P. Thus by (1) we have that M is a p-stable model of Q.
($\Leftarrow$) Let M be a p-stable model of Q. By (1), M is a p-stable model of P. By Lemma 24, M is a MG p-stable model of the abductive logic program of J, $\langle \mathsf{P}, A\rangle$. By Lemma 24, $M \cap A = \emptyset$. Hence by Definition 26, M is an $\odot_{p-stable}$ update model of J.
(**P3. Idempotence**): If P does not have p-stable models, then neither does $P \odot_{p-stable} P$. If P has p-stable models, then $P \cup P$ does, hence by Strong Consistency, $P \cup P \equiv P \odot_{p-stable} P$. Hence in each case $P \odot_{p-stable} P \equiv P$.
(**P4. Weak noninterference**): If each of P_1 and P_2 lacks of p-stable models then the update (in any order) lacks of p-stable models. If P_1 and P_2 have p-stable models, then $P_1 \cup P_2$ does too — because they are defined over disjoint alphabets. By Strong Consistency, $P_1 \cup P_2 \equiv P_1 \odot_{p-stable} P_2$. Also $P_2 \cup P_1 \equiv P_2 \odot_{p-stable} P_1$. Hence, $P_1 \odot_{p-stable} P_2 \equiv P_2 \odot_{p-stable} P_1$.
(**P5. Weak irrelevance of syntax**): Let P, P_1, and P_2 be logic programs under the same language $\mathcal{L}$. Since $P_1 \equiv_{SE} P_2$, then for every program P, $P \cup P_1$ is strongly equivalent to $P \cup P_2$. Thus, $(P \cup A) \cup P_1$ and $(P \cup A) \cup P_2$ have exactly the same p-stable models. Thus, $P \odot_{p-stable} P_1$ and $P \odot_{p-stable} P_2$ have exactly the same EG p-stable models. Therefore, $P \odot_{p-stable} P_1$ and $P \odot_{p-stable} P_2$ have exactly the same MG p-stable models. Hence, $P \odot_{p-stable} P_1 \equiv P \odot_{p-stable} P_2$.
(**P6. Augmented update**): If P_2 does not have p-stable models, neither does $P_1 \odot_{p-stable} P_2$. If P_2 has at least one p-stable model and $P_1 \subseteq P_2$ then, $(P_1 \cup P_2)$ has at least one p-stable model too. By strong consistency $P_1 \odot_{p-stable} P_2 \equiv P_1 \cup P_2$. Hence in each case $P_1 \odot_{p-stable} P_2 \equiv P_2$.
(**P7. Independent parts**): Let $J_1 = (P_1, P_1')$, $J_2 = (P_2, P_2')$ such that $(\mathcal{L}_{J_1} \cap \mathcal{L}_{J_2}) = \emptyset$, and $J = ((P_1 \cup P_2), (P_1' \cup P_2'))$. Let M_1 and M_2 be a $\odot_{p-stable}$ update model of J_1 and a $\odot_{p-stable}$ update model of J_2 respectively. It is clear that M_1 and M_2 are disjoint, since $(\mathcal{L}_{J_1} \cap \mathcal{L}_{J_2}) = \emptyset$. We have to prove that M is a $\odot_{p-stable}$ update model of J iff $M = M_1 \cup M_2$.

($\Rightarrow$) By Definition 26, if M is a $\odot_{p-stable}$ update model of J, then there exists M', a MG p-stable model of $\langle \mathsf{P}, B\rangle$, such that $M = M' \cap \mathcal{L}_J$. Then by Definition 22, there exists Δ, $\Delta \subset B$ such that $\langle M', \Delta\rangle$ is a EG p-stable model of $\langle \mathsf{P}, B\rangle$ and it is minimal. By Definition 21, M' is a p-stable

model of $\mathsf{P} \cup \Delta$.
Moreover, since $(\mathcal{L}_{J_1} \cap \mathcal{L}_{J_2}) = \emptyset$, we can verify the following:
(1) $\mathsf{P} = \mathsf{P}_1 \cup \mathsf{P}_2$[4],
(2)$\Delta = \Delta_1 \cup \Delta_2$ such that $\Delta_1 = \Delta \cap \mathcal{L}_{J_1}$, $\Delta_2 = \Delta \cap \mathcal{L}_{J_2}$ and $\Delta_1 \cap \Delta_2 = \emptyset$,
(3)$M' = M'_1 \cup M'_2$ such that M'_1 is a p-stable model of $\mathsf{P}_1 \cup \Delta_1$ and M'_2 is a p-stable model of $\mathsf{P}_2 \cup \Delta_2$.
Now by Definition 21, $\langle M'_1, \Delta_1 \rangle$ is a MEG p-stable model of $\langle \mathsf{P}_1, B_1 \rangle$ and $\langle M'_2, \Delta_2 \rangle$ is a MGE p-stable model of $\langle \mathsf{P}_2, B_2 \rangle$ where B_1 is the set of abducible atoms of P_1 and B_2 is the set of abducible atoms of P_2.
Finally by Definition 26, $M_1 = M'_1 \cap \mathcal{L}_{J_1}$ and $M_2 = M'_2 \cap \mathcal{L}_{J_2}$ are $\odot_{p-stable}$ update model of J_1 and J_2 respectively.
($\Leftarrow$) This proof is similar to the proof of the first part above. Taking into account that $\mathsf{P} = \mathsf{P}_1 \cup \mathsf{P}_2$; and if $\Delta = \Delta_1 \cup \Delta_2$ then there exists a p-stable model $M' = M'_1 \cup M'_2$ of $\mathsf{P} \cup \Delta$ such that M'_1 as a p-stable model of $\mathsf{P}_1 \cup \Delta_1$ and M'_2 as a p-stable model of $\mathsf{P}_2 \cup \Delta_2$. □

V. Conclusions

Our general schema for defining new update semantics takes as input any basic logic programming semantics S and gives as output a new update semantics. The schema uses minimal generalized S models, where S is any of the logic programming semantics. Each update semantics is associated to an update operator. We also presented properties for the update operator which are valid for the stable and p-stable semantics. The study of those properties for the MM^r semantics as well as the extension of our results to other semantics along with a comparative study of them are topics to be developed in future work.

Acknowledgement: This research has been supported by the Fondo Sectorial SEP-CONACyT, Ciencia Básica Project (Register 101581).

References

[1] J. J. Alferes, F. Banti, A. Brogi, and J. A. Leite. The refined extension principle for semantics of dynamic logic programming. Studia Logica, 79(1):7–32, 2005.

[2] M. Balduccini and M. Gelfond. Logic Programs with Consistency-Restoring Rules. In P. Doherty, J. McCarthy, and M.-A. Williams, editors, International Symposium on Logical Formalization of Commonsense Reasoning, AAAI 2003 Spring Symposium Series, Mar 2003.

[3] P. Baroni, M. Giacomin, and G. Guida. SCC-recursiveness: a general schema for argumentation semantics. Artificial Intelligence, 168:162–210, October 2005.

[4] J. L. Carballido, M. Osorio, and J. Arrazola. Equivalence for the G'$_3$-stable models semantics. J. Applied Logic, 8(1):82–96, 2010.

[5] J. Delgrande, T. Schaub, and H. Tompits. A preference-based framework for updating logic programs. pages 71–83.

[6] J. Dix. A classification theory of semantics of normal logic programs: II. weak properties. Fundam. Inform., 22(3):257–288, 1995.

[7] J. Dix and M. Müller. Partial evaluation and relevance for approximations of stable semantics. In ISMIS, volume 869 of Lecture Notes in Computer Science, pages 511–520. Springer, 1994.

[8] J. Dix, M. Osorio, and C. Zepeda. A general theory of confluent rewriting systems for logic programming and its applications. Ann. Pure Appl. Logic, 108(1-3):153–188, 2001.

[9] T. Eiter, M. Fink, G. Sabbatini, and H. Tompits. Considerations on updates of logic programs. In JELIA '00: Proceedings of the European Workshop on Logics in Artificial Intelligence, pages 2–20, London, UK, 2000. Springer-Verlag.

[10] T. Eiter, M. Fink, G. Sabbatini, and H. Tompits. On properties of update sequences based on causal rejection. Theory and Practice of Logic Programming, 2(6):711–767, 2002.

[11] M. Gelfond and V. Lifschitz. The Stable Model Semantics for Logic Programming. In R. Kowalski and K. Bowen, editors, 5th Conference on Logic Programming, pages 1070–1080. MIT Press, 1988.

[12] A. C. Kakas and P. Mancarella. Generalized stable models: a semantics for abduction. In Proceedings of ECAI-90, pages 385–391. IOS Press, 1990.

[13] J. C. Nieves, M. Osorio, and C. Zepeda. A schema for generating relevant logic programming semantics and its applications in argumentation theory. Accepted in Fundamenta Informatica, 106(2-4):295–319, 2011.

[14] M. Osorio, J. Arrazola, and J. L. Carballido. Logical weak completions of paraconsistent logics. Journal of Logic and Computation, Published on line on May 9, 2008.

[15] M. Osorio and J. L. Carballido. Brief study of G'$_3$ logic. Journal of Applied Non-Classical Logic, 18(4):79–103, 2009.

[16] M. Osorio and V. Cuevas. Updates in answer set programming: An approach based on basic structural properties. Theory and Practice of Logic Programming, 7(04):451–479, July 2007.

[17] M. Osorio, J. A. Navarro, J. Arrazola, and V. Borja. Logics with common weak completions. Journal of Logic and Computation, 16(6):867–890, 2006.

[18] L. M. Pereira and A. M. Pinto. Revised stable models - a semantics for logic programs. In C. Bento, A. Cardoso, and G. Dias, editors, EPIA, volume 3808 of Lecture Notes in Computer Science, pages 29–42. Springer, 2005.

[19] D. van Dalen. Logic and structure. Springer-Verlag, Berlin, 3rd., aumented edition edition, 1994.

[20] F. Zacarias, M. O. Galindo, J. C. A. Guadarrama, and J. Dix. Updates in Answer Set Programming based on structural properties. In Proceedings of the 7th International Symposium on Logical Formalizations of Commonsense Reasoning. Dresden University Technical Report, pages 213–219, Corfu, Greece, May 2005. TU-Dresden, Fakultt Informatik.

[21] C. Zepeda, M. Osorio, J. C. Nieves, C. Solnon, and D. Sol. Applications of preferences using answer set programming. In Answer Set Programming: Advances in Theory and Implementation (ASP 2005), pages 318–332, University of Bath, UK, July 2005.

[4]This is possible if we select the appropriate abducibles from P to define P_1 and P_2 (see Definition 25).