IJCSIS Vol. 12 No. 8, August 2014
ISSN 1947-5500

International Journal of Computer Science & Information Security


Cornell University
Library


Cogprints


Google scholar


.docstoc
find and share professional documents


ScientificCommons

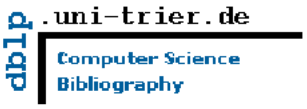

View my documents on
Scribd


BASE
Bielefeld Academic Search Engine


SCIRUS
search engine for science


SciRate.com


CiteSeerx beta


dblp .uni-trier.de
Computer Science
Bibliography


Q·Sensei


DOAJ
DIRECTORY OF
OPEN ACCESS
JOURNALS


EBSCO
HOST


ProQuest

IJCSIS

ISSN (online): 1947-5500

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

CALL FOR PAPERS
International Journal of Computer Science and Information Security (IJCSIS)
January-December 2014 Issues

The topics suggested by this issue can be discussed in term of concepts, surveys, state of the art, research, standards, implementations, running experiments, applications, and industrial case studies. Authors are invited to submit complete unpublished papers, which are not under review in any other conference or journal in the following, but not limited to, topic areas.
See authors guide for manuscript preparation and submission guidelines.

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

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

Context-aware systems
Networking technologies
Security in network, systems, and applications
Evolutionary computation
Industrial systems
Evolutionary computation
Autonomic and autonomous systems
Bio-technologies
Knowledge data systems
Mobile and distance education
Intelligent techniques, logics and systems
Knowledge processing
Information technologies
Internet and web technologies
Digital information processing
Cognitive science and knowledge
Agent-based systems
Mobility and multimedia systems
Systems performance
Networking and telecommunications
Software development and deployment
Knowledge virtualization
Systems and networks on the chip
Knowledge for global defense
Information Systems [IS]
IPv6 Today - Technology and deployment
Modeling
Software Engineering
Optimization
Complexity
Natural Language Processing
Speech Synthesis
Data Mining

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

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

Editorial
Message from Managing Editor

*The **International Journal of Computer Science and Information Security** (IJCSIS) promotes research, review and survey paper publications which offer a significant contribution to the computer science knowledge, and which are of high interest to a wide academic/research/practitioner audience. Coverage extends to all main-stream and state of the art branches of computer science, security and related information technology applications. As a scholarly open access peer-reviewed journal, IJCSIS mission is to provide an outlet for quality research articles. It aims to promote universal access with equal opportunities for international scientific community; to scientific knowledge, and the creation, and dissemination of scientific and technical information.*

*IJCSIS archives all publications in major academic/scientific databases. Indexed by the following International agencies and institutions: Google Scholar, Bielefeld Academic Search Engine (BASE), CiteSeerX, SCIRUS, Cornell's University Library EI, Scopus, DBLP, DOI, ProQuest, EBSCO. Google Scholar reported increased in number cited papers published in IJCSIS (**No. of Cited Papers:524, No. of Citations:1008, Years:5**). Abstracting/indexing, editorial board and other important information are available online on homepage. This journal supports the Open Access policy of distribution of published manuscripts, ensuring "free availability on the public Internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of [published] articles".*

*IJCSIS editorial board, consisting of international experts, ensures a rigorous peer-reviewing process. We look forward to your collaboration. For further questions please do not hesitate to contact us at **ijcsiseditor@gmail.com**.*

,

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

IJCSIS Vol. 12, No. 8, August 2014 Edition

ISSN 1947-5500 © IJCSIS, USA.

Journal Indexed by (among others):

IJCSIS EDITORIAL BOARD

TABLE OF CONTENTS

clustering algorithm to partition the whole input space into several disjoint regions. For each partition, a set of fuzzy rules is extracted based on a f-SVM combining model. Then fuzzy rules sets are used to predict the test data using fuzzy inference algorithms. The performance of the proposed approach is compared with other models using four data sets.

Keywords- Fuzzy rules; Support vector machine - SVM; Self-Organizing Map - SOM; Stock price forecasting; Data-driven model

4. Paper 31071430: Limitations of Current Security Measures to Address Information Leakage Attacks (pp. 26-32)

Omar Hussein, Nermin Hamza, Hesham Hefny
Computer and Information Sciences Department, Institute of Statistical Studies and Research, Cairo University, Egypt

Abstract — Information leakage attacks represent a serious threat for their widespread and devastating effects. Their significance stems from the fact that they are committed by an organization's authorized computer users, and/or processes executing on their behalf. The diverse avenues that could be exploited to carry out such attacks add another barrier towards addressing them. Based on literature review, this paper explores strengths of security measures intended to confront information leakage attacks, and focuses on pinpointing their respective limitations. It demonstrates that only few of them are capable of mitigating such attacks, whereas the rest suffer from conceptual and/or implementation-related limitations that render them vulnerable to circumvention. They are basically prone to high false positive and/or false negative rates, complex to apply, inflexible during execution, suffer from degraded performance, or require hardware modification. Most importantly, neither of them provides a remedy for new undetected malicious software, nor the ever increasing insider threat.

Index Terms—Information Security, Information Leakage, Security Measures, Security Limitations

5. Paper 31071428: Solving the Problem of the K Parameter in the KNN Classifier Using an Ensemble Learning Approach (pp. 33-39)

Ahmad B. A. Hassanat (1), Mohammad Ali Abbadi (2), Ghada Awad Altarawneh (3)
(1, 2) IT Department, (3) Accounting department, Mu'tah University, Mu'tah – Karak, Jordan
Ahmad Ali Alhasanat, College of Business Administration & Economics, Al-Hussein Bin Talal University, Maan, Jordan

Abstract — This paper presents a new solution for choosing the K parameter in the k-nearest neighbor (KNN) algorithm, the solution depending on the idea of ensemble learning, in which a weak KNN classifier is used each time with a different K, starting from one to the square root of the size of the training set. The results of the weak classifiers are combined using the weighted sum rule. The proposed solution was tested and compared to other solutions using a group of experiments in real life problems. The experimental results show that the proposed classifier outperforms the traditional KNN classifier that uses a different number of neighbors, is competitive with other classifiers, and is a promising classifier with strong potential for a wide range of applications.

Keywords- KNN; supervised learning; machine learning; ensemble learning; nearest neighbor;

6. Paper 31071422: Proposing a New Hybrid Approach in Movie Recommender System (pp. 40-45)

Monireh Amini, Department of Computer Engineering, Zanjan Branch, Islamic Azad University, Zanjan, Iran
Mahdi Nasiri, Computer Engineering Department, Iran University of Science and Technology, Tehran, Iran
Mahdi Afzali, Department of Computer Engineering, Zanjan Branch, Islamic Azad University, Zanjan, Iran

Abstract — Due to the unprecedented growth of information, goods and services, a lot of application programs have been created in recent years to help the selection of goods and services to customers. One of the most important application programs are recommender systems that used to things as proposed movies, books, web pages, and E-Business, etc. Most of recommender systems using Collaborative filtering (CF) and or content based filtering (CBF) to provide suggestion for users. In this paper a new approach examined for better assessing the interests of customers. With the understanding of customer behavior, appropriate offer will be provided to customers. In fact, by using a new hybrid approach, weakness of content based filtering and Collaborative filtering methods as much as possible to will be resolve. The results of this paper can be used to keep and attract customers in the institutions or stores that have fixed customers. In this paper, first we reviewing the recommender systems and investigating types of filtering. Then a new hybrid approach by using CF and CBF methods is presented in a movie recommender system. Results are evaluated on movielens valid data, that the results show improvement in the movie recommender system.

Keywords - Hybrid Recommender Systems; Collaborative filtering; Content-based filtering; Hybrid filtering; Spiking Neural Network (SNN); Naive Bayes; E-Business

7. Paper 31071420: Trellis Analysis of Transmission Burst Errors in Viterbi Decoding (pp. 46-53)

Salehe I. Mrutu (1), Anael Sam (2) and Nerey H. Mvungi (3)
(1, 2) School of Computational and Communication Science and Engineering, Nelson Mandela Institution of Science and Technology, Arusha, Tanzania
(3) College of Information and Communication Technologies, University of Dar Es Salaam, Dar Es Salaam, Tanzania

Abstract — The Viterbi decoder is the most favorable solution to the problem of decoding codewords from a convolutional encoder. Viterbi decoder performs exceptionally well when a received codewords block contains single or multiple and scattered errors in a received codewords block. However, the formation of burst errors in data transmission due to high transmission speed and the widely varying error conditions of wireless media in fading channel creates decoding challenge for such conditions which result in unbearable amount of residual errors. By using Viterbi decoders' trellis diagrams, this paper analyses the effects of burst errors to the decoder that lead to residual errors and proposes improvement to the encoding and decoding procedures of the existing (2, 1, 2) binary convolutional encoder. The improved version facilitate effectiveness in the decoder (Viterbi algorithm) in decoding burst errors and hence reduction of residual errors in a poor channel. The proposed enhancements improve the decoder's operational performance by 75 percent. However, the proposed modification reduces the encoder's data transmission rate from 1/2 to 1/6.

Keywords - Locked Convolutional encoder; Bust errors; Residual errors; Non Transmittable Codewords (NTCs); Viterbi Algorithm Decoding

8. Paper 31071436: Towards a Mobile-Based DSS for Smallholder Livestock Keepers: Tanzania as a Case Study (pp. 54-63)

Bernard Mussa & Zaipuna Yonah,*
Computational and Communication Science and Engineering, The Nelson Mandela African Institution of Science and Technology, Arusha, Tanzania.
Charles Tarimo, College of Engineering and Technology, University of Dar Es Salaam, Dar Es Salaam, Tanzania.

Abstract -- Building a useful and responsive Decision Support System (DSS) requires a deep understanding of the pertinent application domain before starting the system design. In this paper we report about an attempt to develop a mobile-based DSS for smallholder livestock keepers with Arusha region as a case study. The objective of the reported study is to provide an information tool for decision making to the smallholder livestock keepers. The development process involved: 1) employing information gathering techniques to understand smallholder livestock keepers' information needs 2) studying the current methods that are used for information flow among livestock

stakeholders. (i.e. smallholder livestock keepers, extension officers and livestock researchers) 3) analysis of the current situation within Arusha: located in the northern parts of Tanzania in terms of mobile phones penetration, with prospects of leveraging the high mobile phone penetration rate for enhanced information sharing among the smallholder livestock keepers and 4) exploration of options for the platform/model to be used for information access and delivery. The outputs of the above four activities were used to inform the requirements elicitation, and design phases of the mobile-based DSS system development. In addition, the mentioned four activities were supplemented by an extensive literature review of related works on requirements engineering in DSS development. It is anticipated that once the system has been developed, it will be of help to livestock keepers, improving farm-level productivity and decision making process. Findings from the study indicate that majority of smallholder livestock keepers in the selected area possess mobile phones and are in need of access to specific information to support their livestock related decision making. However, information access platforms/models that are currently in place do not cater for a satisfactory solution to their needs. Analysis of various options for designing a DSS platform has shown that a model that considers the administrative, organizational structure, as well as roles of relevant stakeholders in the livestock information flow will be useful for the studied context. The proposed Role-based Information Decision Support (RIDS) Model will facilitate data querying, analysis and information delivery based on users' information requirements for the design of the DSS's data marts. This will, in turn, be the basis for implementing a system of information sharing and delivery mechanism that will improve the decision making process and livestock management for smallholder livestock keepers in the studied geographical environment.

Keywords: Decision Support System, Data Mart, Mobile phones, Smallholder livestock keepers.

9. Paper 31071425: Computational Algorithms Based on the Paninian System to Process Euphonic Conjunctions for Word Searches (pp. 64-76)

Rajitha V., Department of Computer Science, Meenakshi College for Women, Chennai, India Research Scholar, Mother Teresa Women's University, Kodaikanal, India
Kasmir Raja S. V., Dean – Research, SRM University, Chennai, India
Meenakshi Lakshmanan, Department of Computer Science, Meenakshi College for Women, Chennai, India

Abstract – Searching for words in Sanskrit E-text is a problem that is accompanied by complexities introduced by features of Sanskrit such as euphonic conjunctions or '*sandhis*'. A word could occur in an E-text in a transformed form owing to the operation of rules of *sandhi*. Simple word search would not yield these transformed forms of the word. Further, there is no search engine in the literature that can comprehensively search for words in Sanskrit E-texts taking euphonic conjunctions into account. This work presents an optimal binary representational schema for letters of the Sanskrit alphabet along with algorithms to efficiently process the *sandhi* rules of Sanskrit grammar. The work further presents an algorithm that uses the *sandhi* processing algorithm to perform a comprehensive word search on E-text.

Keywords – Sanskrit; euphonic conjunction; sandhi; linguistics; Panini; Sanskrit word search; E-text search.

10. Paper 31071411: Classification of Sleep Stages Using Neural Network Based on EEG and EOG signals (pp. 77-79)

Shreya Garg, Vijay Khare
Jaypee Institute of Information Technology, Noida

Abstract — This paper introduces an algorithm for different sleep stages classification. The algorithm consists of wavelet packet transformation (WPT) which is applied to 30 seconds long epochs of EEG and EOG recordings to provide time-frequency information, a feature generator to quantify the information and reduce the data set size, and then artificial neural networks for doing optimal classification. This led to a classification method with efficiency of 90.41 percent.

Keywords-component; Neural Network (NN), Electroencephalograph (EEG), Electrooculorgraph (EOG), Polysomnography (PSG), Wavelet Packet Transform(WPT)

11. Paper 30061432: An Ultra Low Power and High Throughput FPGA Implementation of SHA-1 Hash Algorithm (pp. 80-86)

Shahzad Khan Department of Computer Science Shaheed Benazir Bhutto University(SBBU) Khyber PukhtunKhwa, Pakistan
Zain-ul-Abideen Department of Communications System Engineering School of Electrical Engineering and Computer Sciences(SEECS), NUST, Islamabad, Pakistan
Shahzad Shahid Paracha Department of Communication System Engineering School of Electrical Engineering and Computer Science(SEECS), NUST, Islamabad, Pakistan

Abstract - In this paper, we present a low power and highly parallel SHA-1 architecture which is considered as extremely iterative in nature specifically suitable for power sensitive applications. That is achieved by first identifying non dependent operations among the consecutive iterations of the algorithm and then aligning them for their execution in a highly parallel way. Consequently, when iteration completes, some other iterations also get completed and only a few of their dependent operations are left. By using this approach we were able to perform up to four SHA-1 iterations simultaneously resulting an increase in its throughput approximately by four times. We also explain how our results critically effect in lowering down the power consumption of the design.

Keywords - Cryptography, Hash function, FPGA implementation

12. Paper 30061425: Various Solutions of Black Hole Attack in A mobile Ad Hoc Network (MANET) (pp. 87-92)

Imad I. Saada, Majdi Z. Rashad, Sherihan Abuelenin
Department of Computer Science, Faculty of Computers and Information, Mansoura University, Egypt

Abstract - Mobile ad hoc network (MANET) is a kind of wireless network that has a number of nodes, these nodes are distributed and connected without dependency on any infrastructure. MANET security has been an important issue since many years, many researchers have concerned in the black hole threat which "announce itself that it has a route to the destination in all cases". There are many solutions have been proposed to encounter these threats, the problem is that the security threats still exist because it is not prevented or avoided completely, in addition the performance of MANET is adversely affected by these solutions, the objective is to find out to what degree it is possible to prevent this attack by these solutions without causing negative effect on efficiency of MANET, so this survey may facilitate developing or proposing more compact idea to encounter security threats. This paper discusses many important solutions that work to detect a black hole node by using different strategies. In this paper, a new strategy proposed but still under testing.

Keywords; MANET, Black hole, AODV, LIDBPP, Network security.

13. Paper 30061417: Tools and Techniques for Ontology Interoperability: A Survey (pp. 93-98)

R. Lakshmi Tulasi, Professor& HOD, Department of IT, QISCET, Ongole, India
Dr M. Srinivasa Rao, Professor, Dean CIHL, SIT, JNTUH, Hyderabad, India

Abstract — The idea of the semantic web is to add machine process able information to web-based data in order to realize interoperability. Ontology is a shared conceptualization of knowledge representation of particular domain. These are used for the enhancement of semantic information explicitly. Ontologies play a prominent role in the concept of the semantic web to provide semantic information for assisting communication among heterogeneous information repositories. Ontology Interoperability provides the reusability of ontologies Different domain experts and ontology engineers create different ontologies for the same or similar domain depending on their data modeling requirements. These cause ontology heterogeneity and inconsistency problems. As increasing numbers of ontologies are developed by diverse communities, the demand for rapid ontology mapping is arising. For more better and

precise results ontology mapping is the solution. As their use has increased, providing means of resolving semantic differences has also become very important. Papers on ontology interoperability report the results on different frameworks and this makes their comparison almost impossible. Therefore, the main focus of this paper will be on providing some basics of ontology interoperability and briefly introducing its different approaches. In this paper we survey the approaches that have been proposed for providing interoperability among domain ontologies and its related techniques and tools.

Keywords- Ontology Interoperability; Ontology Mapping; Ontology Alignment; Ontology Merging; Semantic heterogeneity; Semantic web;

14. Paper 31071491: Generic Lightweight Certificate Management Protocol (GLCMP) (pp. 99-105)

Shahzad Khan Department of Computer Science Shaheed Benazir Bhutto University(SBBU) Khyber PukhtunKhwa, Pakistan
Muhammad Asif Department of Communications System Engineering School of Electrical Engineering and Computer Sciences(SEECS), NUST, Islamabad, Pakistan

Abstract - This paper describes a Generic Light Weight Certificate Management Protocol (GLCMP) for handling certificates on mobile devices. Theoretically, various security solutions are designed to protect the valuable information of mobile users. But, its power, memory and processing constraints, high response time and authentication latencies are the main challenges for the researcher to develop and integrate standard security mechanisms in it. It is observed that, most of mobile users are not technical enough to configure security parameters and even already developed libraries do not support extended security features like transparent handling of certificates, verification of identities, and distribution of certificates. In this paper, an innovative and comparatively efficient protocol is designed and implemented. It does not only overcome the shortcoming of the certificate handling in mobile devices but also provides some extended certificate related features like registration, authentication and trust delegation. The designed GLCMP is lightweight because all complex and computation-intensive operations, involved in creation of certificate request in PKCS#10 standard format, are offloaded to a proxy server. It also provides domain based secure registration and verification of the identities without exchanging any confidential information to the proxy servers and even no user's credential is exchanged on network for authentication. After analyzing its performance, we noticed that authentication latency of GLCMP is 0.394 sec which is less than previously proposed protocols like NSI (4.7), PKI (5.01), and PKASSO (5.19 delegation time + 0.082 authentication times). We also formally verified our designed by using Z-Notation Modeling techniques and found that it is protected against man-in-the-middle, replay and impersonation and non-repudiation attacks.

15. Paper 31071421: Efficient RSA Variant for Resource Constrained Environment (pp. 106-112)

Seema Verma, Computer Science Department, Thapar University, Patiala, India
Dr Deepak Garg, Computer Science Department, Thapar University, Patiala, India

Abstract—The work in this paper is concerned with the memory consumption as well as the performance of RSA cryptosystem so that the most popular public key algorithm can be used efficiently in the resource constrained environment also. For this purpose, RSA variant, RC RSA, is proposed which results in low computational cost and low memory consumption. RC RSA is the improvement over dual RSA small e (based on less memory consumption). Mathematically, as compared to Dual RSA, RC RSA results in the increase of decryption speed by a factor of 9 and in implementation roughly by a factor of 6. On the other hand the encryption speed becomes as low as in standard RSA. Besides the computational speed up, RC RSA is proved to be more secure than the Dual RSA scheme.

Keywords- cryptography; encryption; public key; security

16. Paper 31071417: Real Time Recommender System for Music Data (pp. 113-117)

Mrs. Manjula Athani, Prof. Neelam Pathak, Prof. Asif Ullah Khan, Dr. Bhupesh Gour
CSE, T.I.T, RGPV Bhopal , India

Abstract—Recommender system is able to identifying the n-number of users preferences and adaptively recommend music tracks according to user preferences. we are extracting unique feature tempo of each music using Marsyas Tool. Then we are applying BLX- α crossover to a extracted feature of each music track. User favorite and user profiles are included. This system have been emerging as a powerful technique of ecommerce. The majority of existing recommender systems uses an overall rating value on items for evaluating user's preference opinions. Because users might express their opinions based on some specific features of the item, recommender systems could produce recommendations that meet user needs. In this paper we presented a Real time recommender system for music data. Multiuser Real time recommender system combines the two methodologies, the content based filtering technique and the interactive genetic algorithm by providing optimized solution every time and which is based on user's preferences We can also share the favorite songs to other user hence it give better result and better user system.

Keywords-Recommender system, Interactive Genetic algorithm, Content Based filtering BLX- α

17. Paper 31071405: ICT as a Tool For Improving Information Flow Among Livestock Stakeholders. A Case Study of Tanzania (pp. 118-128)

Gladness Mwanga George, Nelson Mandela African Institution of Science and Technology, Arusha Tanzania
Fatma Simba, University of Dar es Salaam,
Zaipuna O.Yonah, Nelson Mandela African Institution of Science and Technology. Arusha Tanzania

Abstract - Traditionally, extension services are used as a means of conveying to rural areas knowledge derived from different studies and surveys. These extension services are an important way to help livestock farmers to improve their lives and production methods, which in turn leads to an increased and improved livestock and livestock products. Nowadays, extension services deliver far more beyond the traditional role. They involve helping farmers to form collaborative groups, solving marketing challenges, etc. This fact has been confirmed by the study reported in this paper. The main objective of the study was to evaluate the current recording system and information flow among livestock stakeholders, and how ICT it has been used as a tool to bridge the gap of information deficit. This paper presents an analysis of data collected from 15 wards from Arumeru district, Arusha region – Northern Tanzania including data from; district council, livestock farmers, extension officers and researchers from three institutions, Livestock Training Agency (LITA), National Artificial Insemination Centre (NAIC) and Nelson Mandela African Institution of science and Technology (NM-AIST). The results reveal that the current recording system is poor and there is a gap in the flow of information among the stakeholders of livestock farming. This gap has significantly contributed to the deterioration of access to extension services to livestock farmers. Along with other factors, this gap is also attributed to the researchers, who publish their research findings through various online journals. The gap signifies that it has been hard for extension officers to fulfil their roles due to a large number of farmers they serve and the challenge of traveling long distances to deliver such services to farmers. Based on the results of this analysis, the paper concludes with a set of identified approaches on how ICT can play a part in minimizing the gap that has been found for increasing the efficiency of the extension services in reaching livestock farmers.

Key words: ICT, information, Extension services, livestock stakeholders, record keeping.

Confidential Algorithm for Golden Cryptography Using Haar Wavelet

Marghny H. Mohamed
Computer Science Department
Faculty of Computers and Information
Assuit University
Egypt

Yousef B. Mahdy
Computer Science Department
Faculty of Computers and Information
Assuit University
Egypt

Wafaa Abd El-Wahed Shaban
Information Systems Department
Faculty of Computers and Information
Assuit University
Egypt

***Abstract*— One of the most important consideration techniques when one want to solve the protecting of digital signal is the golden matrix. The golden matrices can be used for creation of a new kind of cryptography called the golden cryptography. Many research papers have proved that the method is very fast and simple for technical realization and can be used for cryptographic protection of digital signals. In this paper, we introduce a technique of encryption based on combination of haar wavelet and golden matrix. These combinations carry out after compression data by adaptive Huffman code to reduce data size and remove redundant data. This process will provide multi-security services. In addition Message Authentication Code (MAC) technique can be used to provide authentication and the integrity of this scheme. The proposed scheme is accomplished through five stages, the compression data, key generation, encryption stage, the decryption stage and decompression at communication ends.**

Keywords: Cryptography, Golden matrix, Adaptive Huffman Compression, Haar wavelet, Message Authentication Code(MAC)

I. INTRODUCTION

The main challenge in data communication is focused on how to keep data secure against unlawful interference. One of the common serious attacks which threaten data security today is: intercepted; which occurs when an unauthorized party can access to read protected file and modify data. Many papers try to improve golden cryptography to solve this challenge [7, 11, 12, 13, 15]. Cryptosystems rely on the assumption that a number of mathematical problems are computationally intractable in the sense that they cannot be solved in polynomial time.

The simplicity and beauty of Fibonacci numbers have been motivated to develop matrix cryptosystems, which are useful in digital communications, i.e., digital TV, digital telephony, digital measurement, etc. One of such cryptosystems, called the golden cryptography based on the golden matrices, which are a generalization of Fibonacci Q-matrices for continuous domain, was introduced by Stakhov [13]. Any cryptosystem is considered to be secure if it is resistible against different types of cryptanalytic attacks such as the ciphertext-only attack, the known-plaintext attack and the chosen-plaintext (chosen ciphertext) attack, etc. In case of chosen plaintext attack, the cryptanalyst can obtain the ciphertexts corresponding to an arbitrary set of plaintexts of his own choosing. Unfortunately, Rey and Sanchez [9] showed that the cryptosystem proposed [13] is not secure against chosen plaintext attack, where the secret key can be obtained easily. Another interesting cryptosystem based on Hadamard product of golden matrices was introduced by Nally [11]. There are also other simple cryptographic methods [7, 12, 15] based on extensions of golden matrices. M.Tahghighi, et al., proved that these methods are also insecure against chosen-plaintext attack [1]. So in this paper, we will try to solve the problem by proposing an improved version of golden cryptography by using Haar wavelet for golden matrix (Fibonacci Numbers, ELC Numbers and Lucas Numbers). This leads that the proposed approach own the powerful properties of the haar wavelet such as orthonormality, compact support, varying degrees of smoothness, localization both in time or space and scale (frequency), and fast implementation. In addition, one of the key advantages of wavelets is the ability to adapt to the features of a function such as discontinuities and varying frequency behavior [19, 21]. Traditional cryptographic algorithms, such as DES, AES, RSA, etc. [16, 20] send the ciphertext over the cyberspace while keeping a secret part (i.e. key) shared, which tends to be dangerous, as any intruder can get the encrypted message and apply his own cryptanalysis techniques, this means when data travel over the network even though it is hidden more attacks could be applied to the cipher message trying to get full or partial information from the message. In our scheme, the data sent over the communication channel are not the original encrypted message, but this is the compressed one and the encipher matrix generated in the sender side and the decipher matrix generated in receiver side while keeping secret part (i.e. N recurrence sequences , number of haar wavelet level and type of recurrence matrix) shared. Also, sending the HMAC of the compressed data along with the cipher enables the receiver party to verify the sender identity and message integrity. Thus, our model is carried out by several mechanisms like adaptive Huffman coding, Recurrence relations, Haar wavelet, Hash based Message Authentication Code (HMAC) to build the encryption phase.

A. Adaptive Huffman Coding

Huffman coding needs some knowledge of the probabilities of the source sequence. If this information is unavailable, compressing the file requires two passes: the statistics are

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

collected in the first pass, and the source is encoded in the second pass. In adaptive huffman coding convert this algorithm into a one-pass procedure, neither transmitter nor receiver knows anything about the source sequence at the start of transmission.

Both the transmitter and the receiver build tree consisting of a single node that corresponds to all symbols not yet transmitted (NYT) and has a weight of zero. As transmission progresses, nodes corresponding to symbols transmitted will be added to the tree, and the tree is reconfigured using an update procedure. Both transmitter and receiver start with the same tree structure. The updating procedure used by both transmitter and receiver is identical. Therefore, the encoding and decoding processes remain synchronized [17].

1) Update Procedure: The update procedure requires that the nodes be in a fixed order. This ordering is preserved by numbering the nodes. Figure 1 is a flowchart of the updating procedure [17].

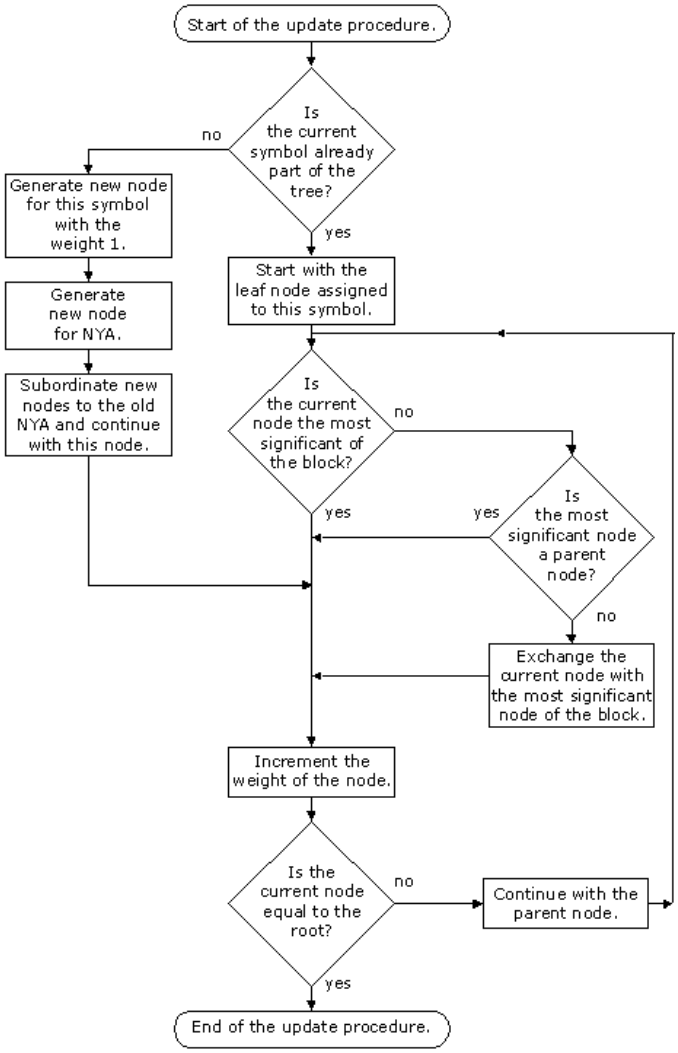


Figure 1. Update procedure for the adaptive Huffman coding algorithm.

2) Encoding Procedure: The flowchart for the encoding procedure is shown in Figure 2 [17].

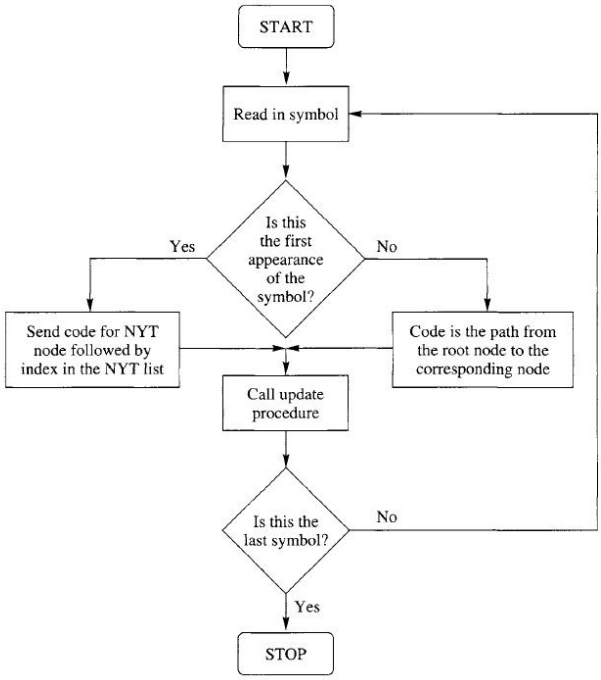


Figure 2. Flowchart off the encoding procedure

B. RECURRENCE RELATIONS

Recurrence relation is useful in certain counting problems like Fibonacci numbers, Lucas and ELC. A recurrence relation relates the nth element of a sequence to its predecessors. Recurrence relations are related to recursive algorithms. A recursive relation for the sequence a_0; a_1; a_2; …… is an equation that relates a_n to certain of its preceding terms a_0; a_1; a_2; ….. a_{n-1}.

Initial conditions for the sequence a_0; a_1; a_2….. are explicitly given values for a finite number of the terms of the sequence.

In this section recurrence relations Fibonacci, Lucas and ELC numbers were presented and their application to cryptography is examined [7, 12-13, 22-24].

1) Fibonacci numbers: Fibonacci numbers are given by the following recurrence relation [22-23]

$$F_{n+1} = F_n + F_{n-1} \qquad (1)$$

With the initial conditions

$$F_1 = F_2 = 1 \qquad (2)$$

A square matrix (2 × 2) as shown below was introduced in [22]

$$Q = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \qquad (3)$$

The following property of the nth power of the Q-matrix was proved

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

$$Q^n = \begin{pmatrix} F_{n+1} & F_n \\ F_n & F_{n-1} \end{pmatrix} \quad (4)$$

Where n = 0, ±1; ±2; ±3; . . ., F_{n-1}; F_n; F_{n+1} are Fibonacci numbers.
Hence, the inverse of matrices Qn is

$$Q^{-n} = \begin{pmatrix} +F_{n-1}/(-1)^n & -F_n/(-1)^n \\ -F_n/(-1)^n & +F_{n+1}/(-1)^n \end{pmatrix} \quad (5)$$

The generalized Fibonacci matrix Qp is defined by

$$Q_p = \begin{pmatrix} 1 & 1 & 0 & . & . & . & . & . & 0 \\ 0 & 0 & 1 & . & . & . & . & . & 0 \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & . & . & 0 & 1 \\ 1 & 0 & 0 & . & . & . & . & 0 & 0 \end{pmatrix} \quad (6)$$

Note that the Q_p-matrix is a square (p + 1) × (p + 1) matrix.
For p = 0, 1, 2, 3 ... the Q_p-matrices have the following forms, respectively:

$$Q_0 = \begin{pmatrix} 1 \end{pmatrix} \quad Q_1 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \quad (7)$$

$$Q_2 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \quad Q_3 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix} \quad (8)$$

In general the nth power of the Q_p matrix

$$Q_p^n = \begin{pmatrix} F_p(n+1) & F_p(n) & \cdots & F_p(n-p+2) & F_p(n-p+1) \\ F_p(n-p+1) & F_p(n-p) & \cdots & F_p(n-2p+2) & F_p(n-2p+1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ F_p(n-1) & F_p(n-2) & \cdots & F_p(n-p) & F_p(n-p-1) \\ F_p(n) & F_p(n-1) & \cdots & F_p(n-p+1) & F_p(n-p) \end{pmatrix}$$

2) Lucas numbers: The sequence of Lucas numbers L_k is defined by the second-order linear recurrence formula and initial terms

$$L_{k+1} = L_k + L_{k-1}, L_0 = 2, L_1 = 1 \quad (9)$$

The proposed matrix using Lucas recursion

$$L^n = \begin{pmatrix} L_{n+1} & L_n \\ L_n & L_{n-1} \end{pmatrix} \quad (10)$$

$$L^1 = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} \quad (11)$$

The inverse of matrices L^n is

$$L^{-n} = \begin{pmatrix} +L_{n-1}/((-1)^n * 4) & -L_n/((-1)^n * 4) \\ -L_n/((-1)^n * 4) & +L_{n+1}/((-1)^n * 4) \end{pmatrix} \quad (12)$$

The other explicit forms of Ln can be obtained recursively same as Q^n

2) ELC numbers: ELC numbers are given by the following recurrence relation $E_{n+1} = E_n + E_{n-1}$, with condition $E_0 = 8$ and $E_1 = 14$. The golden matrix using ELC recursion is proposed as follows.

$$E^n = \begin{pmatrix} E_{n+1} & E_n \\ E_n & E_{n-1} \end{pmatrix} \quad (13)$$

Where n = 0, 1, 2, 3, the inverse of matrices E^n is

$$E^{-n} = \begin{pmatrix} +E_{n-1}/((-1)^n * 20) & -E_n/((-1)^n * 20) \\ -E_n/((-1)^n * 20) & +E_{n+1}/((-1)^n * 20) \end{pmatrix} \quad (14)$$

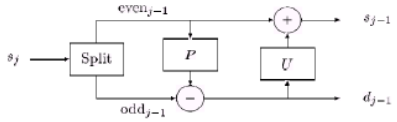
The other explicit forms of E^n can be obtained recursively same as Q^n

C. Haar Wavelet Transform

Haar wavelet is the simplest wavelet. Haar transform or Haar wavelet transform has been used as an earliest example for orthonormal wavelet transform with compact support [19, 21]. The Haar wavelet transform is the first known wavelet and was proposed in 1909 by Alfred Haar. The Haar wavelet transform has a number of advantages:

- It is conceptually simple.
- It is fast.
- It is memory efficient, since it can be calculated in place without a temporary Array.

1) Procedure for Haar Wavelet Transform: To calculate the Haar transform of an array of n samples [18]:



1) *Split*: divide the input data into:
- Even indexed samples S_n.
- Odd indexed samples S_{n+1}. Lazy wavelet transform

2) Predict: the odd elements from the even elements-output detail.

$$d_{j-1} = s_j[2n+1] - s_j[2n]. \quad (15)$$

In general:

$$d_{j-1} = odd_{j-1} - P(even_{j-1}). \quad (16)$$

3) Update:
- Follows the predict phase.
- The approximations S_{n-1}(the signal for next step) should maintain the average of the original signal S_n.

$$s_{j-1}[n] = s_j[2n] + d_{j-1}[n]/2. \quad (17)$$

In general:

$$s_{j-1} = even_{j-1} + U(d_{j-1}). \quad (18)$$

D. Hash-*based* Message Authentication Code (HMAC)

The purpose of an MAC is to authenticate both the source of a message and its integrity without the use of any additional mechanisms. HMACs have two functionally distinct parameters, a message input and a secret key known only to the message originator and intended receiver(s). Additional applications of keyed-hash functions include their use in challenge-response identification protocols for computing responses, which are a function of both a secret key and a challenge message. An HMAC is used by the message sender to produce a value (the MAC) that is formed by condensing the secret key and the message input. The MAC is typically sent to the message receiver along with the message [8]. The receiver computes the MAC on the received message using the same key and HMAC function as was used by the sender, and compares the result computed with the received MAC. If the two values match, the message has been correctly received, and the receiver assures us that the sender is a member of the community of users that share the key, as shown in Figure 3.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

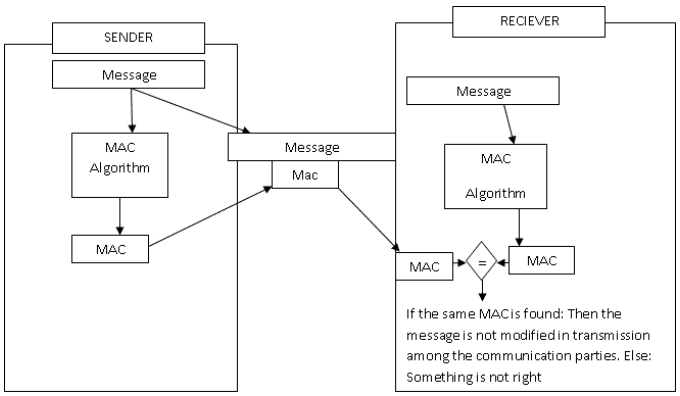


Figure 3. The use of MAC

HMAC-SHA-256+ are secret key algorithms. While no fixed key length is specified in [HMAC], key lengths less than the output length decrease security strength, and keys longer than the output length do not significantly increase security strength [14].

The rest of this paper is organized as following, the proposed scheme is presented in section 2, in section 3 the security analysis is introduced, experimental results is presented in section 4, finally conclusions are provided in section 5.

II. THE PROPOSED SCHEME

This section examines improving golden cryptography to solve insecurity against the chosen-plaintext attack for golden cryptography by using haar wavelet transform, in terms of the problems the secret key can be obtained for example, A.Stakhov suggested"new kind of cryptography system" in [13] but Rey and Sanchez showed that this cryptosystem is not secure against chosen plaintext attack [9], which they let pairs of plaintext M1, M2, M3 and M4, which

$$M1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, M2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$
$$, M3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, M4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

And the enciphering matrix is Q^{2x}

$$Q^{2x} = \begin{pmatrix} cFs(2x+1) & sFs(2x) \\ sFs(2x) & cFs(2x1) \end{pmatrix}$$

For a more detailed description of such functions we refer the reader to [1, 9]. Using simple calculus shows, the real value for x is:

$$x = \frac{1}{2}\log_{\tau}\left(\frac{\sqrt{k_1\sqrt{5}+\sqrt{5k_1^2+4}}}{2}\right), \quad (19)$$

where $k_1 \in R$ can be obtained from the equation

$$z = \frac{k_1\sqrt{5} \pm \sqrt{5k_1^2+4}}{2}, \quad (20)$$

Where $z = \tau^{2x}$. τ is golden proportion, thus the secret key x is obtained [9]. By the similar calculations M.Tahghighi, et al. proved that these methods [7, 12, and 15] insecure against chosenplaintext attack [1], which the secret key x of the Hadamard product of golden matrices can be obtained by:

$$x = \frac{1}{4}\log_{\tau}\left(\frac{2-2.5(t_1-t_2) \pm \sqrt{6.25(t_1-t_2)^2+5(t_1-t_2)}}{2}\right), \quad (21)$$

known real variables t_1 and t_2, we obtain the system

$$Fs(2x+1)cF(2x1) - [sFs(2x)]^2 = t_1, \quad (22)$$

$$cFs(2x+1)cF(2x-1) + [sFs(2x)]^2 = t_2, \quad (23)$$

of non-linear equations. By the definition of $s\mathrm{F}(x)$ and $c\mathrm{F}s(x)$,

$$\frac{(\tau^{2x+1}+\tau^{-2x-1})}{5} - \frac{(\tau^{2x}-\tau^{-2x})}{5} = t_1, \quad (24)$$

$$\frac{(\tau^{2x+1}+\tau^{-2x-1})}{5} + \frac{(\tau^{2x}-\tau^{-2x})}{5} = t_2, \quad (25)$$

and the secret key $\{k, x\}$ is obtained of the generalized golden cryptographic method it is easy to see that

$$k = \frac{\sigma_k^2 - 1}{\sigma_k}, \quad (26)$$

we can calculate the value of x, i.e.,

$$x = \frac{1}{2}\log\sigma_k(t_1\sigma_k + t3) \ (or\ x = \frac{1}{2}\log\sigma_k(t_1\sigma_k^{-1} + t_2)), \quad (27)$$

they are very simple and it is very easy to show their insecurity against the chosen-plaintext attack [1].

The proposed scheme can be summarized in the following stages: At *sender side* some stages must be done:

- Compression data

1) Map each character in plaintext into its corresponding ASCII code, M'=ASCII(M).

2) Compression M' by using adaptive Huffman Coding and generate compressed data CM.

- Producing MAC Message

3) Generate key by using any encryption algorithm (e.g DES, TripleDES, AES) to compute a MAC over the compressed data message CM using the HMAC function.

- Encryption Stage

1) Input a cryptographic key, K=n, r.

2) Construct the corresponding "Golden matrix" G Matrix Depending on r equivalent $Q^n{}_p$ or $L^n{}_p$ or $E^n{}_p$ (calculate p depended by message size p=ceil($\sqrt{M}$).

3) Compute key encryption matrix E, where equivalent Haar wavelet matrix from G Matrix according to l this create random and add another number matrix.

4) Break up CM into CG groups (each group contains rowmatrix2 elements) and from each on a square matrix.

*5) For each group, compute the corresponding CipherText where Ci = CG*E.*

6) Collect Ci and send its.

Algorithm.1 shows the steps at sender side.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

```
Data: M, n, l, r
where M is Plaintext, n is used as short session key(one
time pad) and l is level, r type of Recurrence matrix
Result: chiphertext
initialization;
/* Map each character in plaintext into
   its corresponding ASCII code.          */
M' ⟵ ASCII(M);
/* Compression M' by using adaptive
   Huffman Coding and generate
   compressed data CM .                   */
CM ⟵ adaptiveHuffmanCoding(M');
/* Producing MAC Message :Generate key
   by using any encryption algorithm
   (e.g DES, TripleDES, AES) to compute
   a MAC over the compressed data
   message CM using the HMAC function.
   */
ComputeHash(CM);
/* Generate Encryption Key                */
/* Golden matrix G Matrix Depending on
   r equivalent Q_p^n or L_p^n or E_p^n   */
G ⟵ GeneralGoldenMatrix(n, p);
/* Compute key encryption matrix E,
   where equivalent Haar wavelet matrix
   from G Matrix according to l and add
   another number matrix.                 */
E ⟵
WaveletTransform(G, l) + anothernumbermatrix;
while not at end of CM do
    if
    (CM.Length − CM.Position) > (E.Row ∗ E.Col)
    then
        bytesToRead = (E.Row ∗ E.Col);
    else
        bytesToRead = (CM.Length − CM.Position);
    end
    CM.Read(buf, 0, bytesToRead);
    while not at end of buf do
        index=0 ;
        for i ← 0 to E.Row do
            for j ← 0 to E.Col do
                if index >= buf.Length then
                    m2[i, j] ⟵ −1;
                else
                    m2[i, j] ⟵ buf[index];
                    index=index + 1;
                end
            end
        end
        m3 ⟵ Matrix.Multiply(m2, E);
    end
    /* Collect m3 in chiphertext.         */
    chiphertext ⟵ m3;
end
```

Algorithm 1: Algorithm steps at sender side

At the *receiver side* another steps are done to decrypt the ciphertext and retrieve the original message, in addition to ensuring authenticity and message integrity according to the following stages:

- Decryption Stage

1) Compute inverse of E matrix.

2) Break up the ciphertext into CG groups and a square matrix for each block.

*3) For each group, compute CG =Ci *E−1.*

4) Collect CG, where Compressed message.

5) Decompress CG to Original message M'.

6) Map each ASCII in M' to corresponding character M.

- Verification Stage

1) Compute MAC value of the obtained Compressed message (Obt-MAC).

2) Compare the obtained MAC value with the MAC value of the constructed message Obt-MAC, if the matching obtained (MAC = Obt-MAC). This indicates that the message is not modified in transmission among the communication parties.

Algorithm.2, shows the steps at receiver side.

```
Data: CM
where CM is Chiphertext
Result: PlainText
initialization;
/* Generate Decryption Key                */
/* Compute (IE) inverse of E matrix
   where equivalent Haar wavelet matrix
   from G Matrix.                         */
IE ⟵ InverseE();
while not at end of CM do
    Cipher ⟵ CM.read();
    while not at end of Cipher do
        index=0 ;
        for i ← 0 to IE.Row do
            for j ← 0 to IE.Col do
                m2[i, j] ⟵ Cipher[index];
                index=index + 1 ;
            end
        end
        m3 ⟵ Matrix.Multiply(m2, IE);
    end
    /* Collect m3 in Plaintext.           */
    CPlaintext ⟵ m3;
end
/* DeCompression M' by using adaptive
   Huffman Coding and generate
   compressed data CM .                   */
Plaintext ⟵
adaptiveHuffmanCodingDecode(CPlaintext);
/* Compare the obtained MAC value with
   the MAC value of the constructed
   message Obt−MAC, if the matching
   obtained (MAC = Obt−MAC). This
   indicates that the message is not
   modified in transmission among the
   communication parties.                 */
Obt − MAC ⟵ ComputeHash(CPlaintext);
if Obt-MAC.SequenceEqual(MAC) then
    "Equality" ;
else
    "Data attack" ;
end
```

Algorithm 2: Algorithm steps at receiver side

III. EXPERIMENTAL RESULTS

In order to evaluate the effectiveness of the proposed scheme, the following experiment has been conducted to measure the level of confusion and diffusion, by comparing plain to cipher the relationship as a metric model for security. These simulation experiments have been done on a sentence M representing the original message: M ="Cryptographist is the science of overt secret writing", to encrypt this message by the proposed model. Suppose k= 5, 2. The contrast between plaintext and ciphertext is demonstrated in Figure.4.

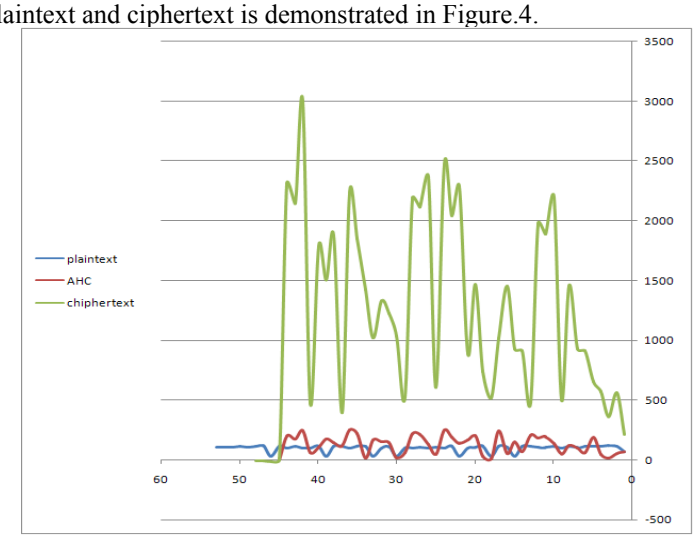


Figure 4. The Contrast between Plaintext and Ciphertext

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

From the resulted ciphertext we can clearly notice that for each character on the original message there is a different value appeared in the ciphertext, and there is no direct relationship between the plaintext and the cipher text. The benefits of use adaptive huffman code are reducing data and removing redundant data. This indicates that the proposed model has a high confusion because the relationship between the input (key) and the output (message) is nonlinear. We observed that the message has some repeated characters such as character"e" for example (repeated six times), and every time the resulted cipher is different from the other, the repeated values disappeared on the resulted ciphertext. Figure .5 shows the distribution of the"e" character in the ciphertext. This indicates that the proposed model provides a high-level of diffusion.

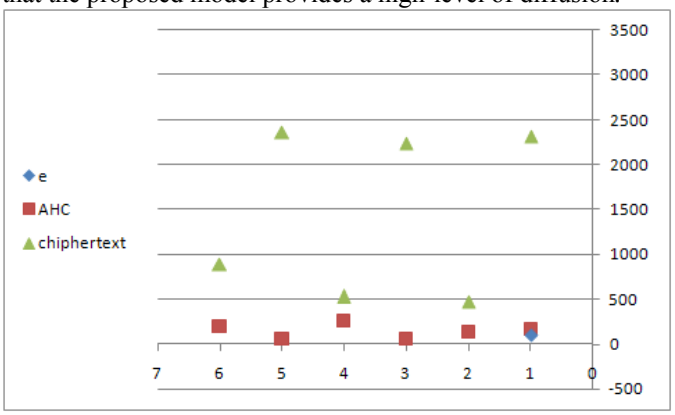


Figure 5. Distribution of character 'e' on ciphertext

A similar experiment has also been conducted to a sentence consisting of consecutive m's as a plaintext with the length = 10 on this message M1 = "mmmnmmmomm". As we can see clearly, the resulted ciphertext is completely different from the ciphertext although the M1 character is eight times in a sentence. The contrast between plaintext and ciphertext is demonstrated in Figure.6.

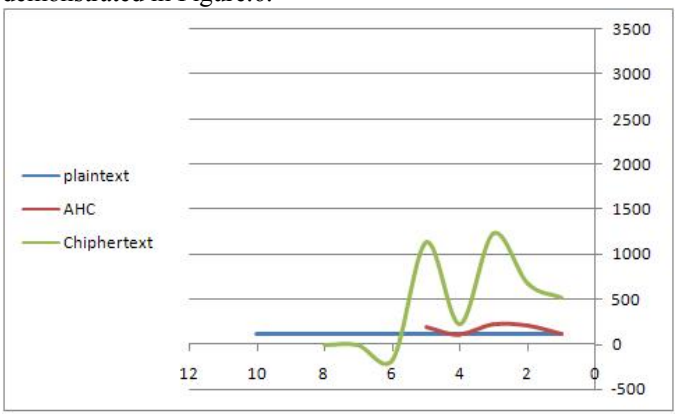


Figure 6. The Contrast between Plaintext and Ciphertext

To confirm our results one more experiment is conducted. We encrypted another message similar to the previous one using the same key used before, to see what happens when two very similar texts are encrypted under the same key. These simulation experiments have been done on a sentence M2 representing the original message: M2 ="meet me after party meet me after party", again we can see, the resulted cipher is totally different from the previous experiment as shown in Figure.7, although it is the same message and encrypted under similar key.

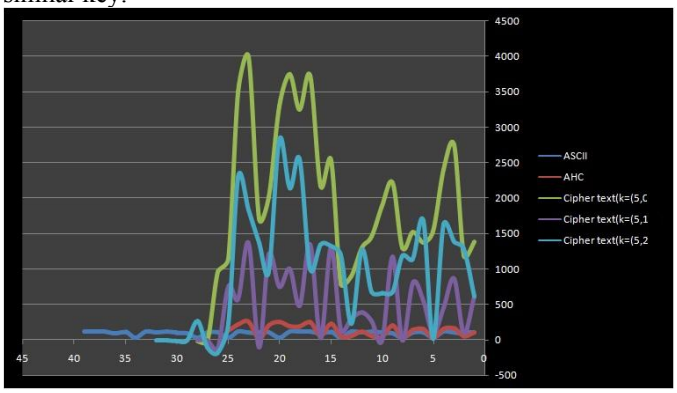


Figure 7. The Contrast between the same message and its Ciphertext under similar keys

In order to evaluate the effectiveness of modify our proposed scheme to increase security and increase confusion, the following experiment has been conducted to measure the level of confusion and diffusion, by comparing plain to cipher relationship as a metric model for security about simulation experiments have been done on a sentence M2 representing the original message: M2 ="meet me after party meet me after party". This experiment has been conducted to measure the level of confusion in modifying our scheme by generating matrix key different size depending on level random generate and p about Q^n_p or L^n_p or E^n_p , this show in Figure .8.

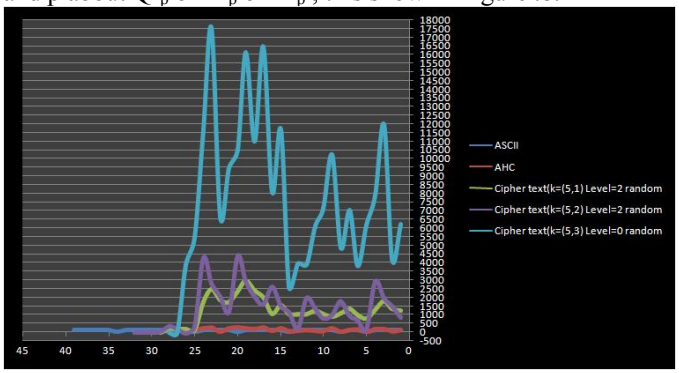


Figure 8. The Contrast between Plaintext and Ciphertexts

In Figure.9 Comparison between schema and modify schema, we notice increase the level of confusion from the previous scheme between CipherTexts

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

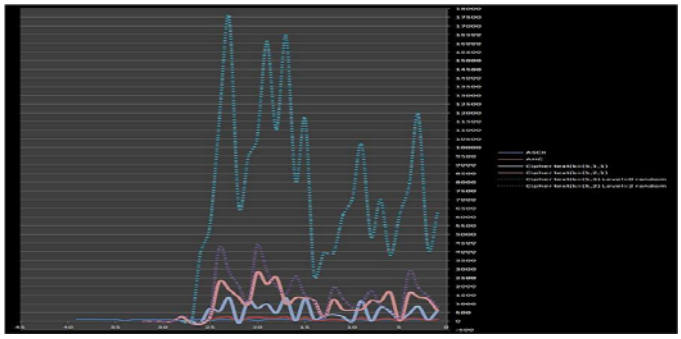


Figure 9. Comparison between schema and modify schema

Even under this schema and modify schema, no relation between plaintext and ciphertext can be noticed, and the distribution of ciphertext is random. When we increase the message length with repetition, no relation could be noticed between ciphertext and plaintext indicates the strength of this scheme against partially known plaintext attack. This confirms what we mentioned the confusion and diffusion properties are provided by the proposed scheme. The performance of the secret key algorithms has been compared on different data, by using input file data of varying sizes and formats.

IV. Security Analysis

Some security analysis has been performed on the proposed encryption scheme [2-6 and 10], such as:

- *Known-plaintext attack*: Suppose the intruder knows some pairs of ciphertexts and corresponding plaintexts, here his goal is to reveal the shared data (keys), to use it in future to decipher other ciphertext. The intruder will then have to search in a semi-impossible search space. Consider the Haar wavelet in the proposed scheme; the attacker must firstly construct the set of possible key space. If we suppose that the key size of the Haar wavelet is Z*Z, where Z is equal 2^{level} if the number of golden row matrix is already a 2^{level} otherwise Z is the next bigger number that's a 2^{level}, where level and p is depending on the data size used. An example to illustrate the enciphering matrix: Let K=1, 1 where N = 1, R = 1, P = 0, L = 1

1) Generate recursion matrix
$Q_0^1 = (\ 1\ )$.
2) Add padding to original matrix and calculate extra row, col
$E = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.
3) Row transform
$E = \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix} \rightarrow E = \begin{pmatrix} 0.5 & -1 \\ 0 & 0 \end{pmatrix}$.
4) Column transform
$E = \begin{pmatrix} 0.5 & -1 \\ -0.5 & 0 \end{pmatrix} \rightarrow E = \begin{pmatrix} 0.25 & -1 \\ -0.5 & 0 \end{pmatrix}$
$\rightarrow E = \begin{pmatrix} 0.25 & -1 \\ -0.5 & 1 \end{pmatrix}$.
5) Cryptography key
$E = \begin{pmatrix} 0.25 & -0.5 \\ -0.5 & 1 \end{pmatrix}$.

When change L to 2

$$E = \begin{pmatrix} 0.0625 & -0.1250 & -0.5000 & 0.0000 \\ -0.1250 & 0.2500 & 0.0000 & 0.0000 \\ -0.5000 & 0.0000 & 1.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{pmatrix}.$$

To apply scheme on this matrix, we can get more than enciphering matrices using the same key by adding random matrices to covering the 0's in E, and this example shows that the enciphering matrices size increase with an increasing level of haar wavelet, where this is exponential increasing. In addition to the one way property of her wavelet is nonlinear makes computing the key of it (inputs) difficult enough even if the intruder knows the ciphertext and the plaintext, which is not the original text, but this is the compressed one.

- *Ciphertext-only attack:* Suppose the intruder can eavesdrop the ciphertext in transmit, his goal here is to reveal the keys (inputs), or the corresponding plaintext (outputs), then the intruder must search in the key space of inputs, and in the key space of inputs is (z*z) this is two dimensions matrix depending on the level of wavelet transformation, p of recurrence matrix and type of recurrence use (Fibonacci, ELC, Lucas), then the intruder must search in the following key space: (z*z) and random matrix, the bigger Z and random matrix are the larger the key space which search on it. Also the plaintext is compressed data the time complexity of an adaptive Huffman encoding is linear: $N[\Sigma + log(2\Sigma - 1)] + Sn$ where N is the total number of input symbols, $\sum$ is the current number of unique symbols, and S is the time required, if necessary, to rebalance the tree [26, 27], and encryption of compressed data each block consists of the $T_e = G(Z^3 \triangle t_m + Z^2(Z\text{-}1)\triangle t_a)$ where $\triangle t_m$ is a time of one multiplication and $\triangle t_a$ is a time of one addition. The data complexity about encryption is $O(N \log[\Sigma]) + O(2^{level}) + O(Z^3)$ Due to this fact the cipher difficult to be broken.
- *Confusion and diffusion:* Confusion and diffusion are two basic design criteria for encryption algorithms [25]. Diffusion means spreading out the influence of a single plaintext symbol over many ciphertext symbols so as to hide the statistical structure of the plaintext. Confusion means the use of transformations to complicate the dependence of the statistics of ciphertext on that of the plaintext. The proposed cryptosystem has a high confusion and diffusion properties, which makes the cryptosystem of high key sensitivity and plaintext sensitivity, and this of high computing security. On the other hand, the mapping function of the used Golden cryptography is a nonlinear function which makes the relationship between the plaintext, key and ciphertext nonlinear. This property complicates possibility of retrieving one of them even if the others were known [8].
- *Statistical analysis:* Correlation Coefficient Analysis and t-test Statistical analysis such as correlation coefficient factor is used to measure the relationship between two variables. This factor examines the proposed encryption algorithm which strongly resists statistical attacks. Therefore, ciphertext must be completely different from the plaintext. The paired t-test and correlation use the same type of data; it is

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

easy to confuse the two techniques. The paired t-test is used to test for differences in the mean values of each variable, while correlation shows associations between the pairs of values. If the correlation coefficient equals one, that means the plaintext and its encryption is identical. If the correlation coefficient equals zero, that means the ciphertext is completely different from the plaintext (i.e. good encryption). If the correlation coefficient equals minus one that means the ciphertext is the negative of the plaintext. So, success of the encryption process means smaller values of the correlation coefficient. The experimental results, the correlation coefficient value and Paired t-test of the proposed encryption algorithm is:

TABLE I. THE CORRELATION AND PAIRED T-TEST

The Correlation and paired t-test from encrypted msg2		
meet me after the party meet me after the party		
Ciphertext	***Correlation***	***Paired t-test***
C1m1	0.401154711	0.0001
C1m2	0.24715579	0.0001
C2m1	0.401826534	0.0001
C2m2	0.286313711	0.0001
C3m1	0.376013828	0.0001
C3m2	0.152848002	0.0001
Rijndael	0.260303958	0.08258
DES	0.446170523	0.0040
TripleDES	-0.400377119	0.0775

TABLE II. THE UNPAIRED T-TEST

The Unpaired t-test from encrypted msg2			
meet me after the party meet me after the party			
	C1m2	***C2m2***	***C3m2***
C1m1	0.0006	0.0001	0.0002
C2m1	0.0006	0.0001	0.0002
C3m1	0.0007	0.0001	0.0002

Where C1m1, C1m2, C2m1, C2m2, C3m1 and C3m2 are chiphertexts under the same key and different random matrix. In case of chiphertext the values of correlation coefficient between 0.152848002 and 0.401154711, which means that the schema is uncorrelated. The values of t-test show that this difference is considered an extremely statistically significant; this means that the proposed encryption algorithm has a strong security.

V. CONCLUSION

This paper is proposing a cryptosystem based on hybrid approaches proposed. The proposed cryptosystem provided multi-security services such as confidentiality, authentication, and integrity, which there are important security services in most applications. To serve data confidentiality by using a technique of encryption based on combination of haar wavelet and golden matrix. These combinations are carried out after compression data by adaptive Huffman code to reduce data size, remove redundant data and consider this initial encryption to the data because it becomes more sensitive to transmission errors or any change in data by an intruder. MAC technique produced the digital signature of the scheme for providing the other mentioned security services. In this scheme the digital signature firstly produced by computing the MAC of the compressed message, then signing it by the sender's private key generated by any encryption algorithm. Finally, the signed MAC and cipher will be sending to cyberspace. At the receiver end, after the decryption is done, the digital signature about compressed data can be used to verify the integrity of the message and the authentication of the sender then decompress the data to get at the original data. The experimental results indicate that the proposed cryptosystem has a high confusion and diffusion properties, it has high security and it is suitable for secure communications.

REFERENCES

[1] M.Tahghighi, S. Turaev, A.Jaafar, R. Mahmod and M.Md.Said, On the Security of Golden Cryptosystems, Int. J.Contemp. Math Sciences, Vol.7, 2012,p. 327 - 335.

[2] H. Marghny Mohamed, M. Naziha Al-Aidroos and A. Mohamed Bamatraf, Data Hiding Technique Based on LSB Matching towards High Imperceptibility, MIS Review 2012 Department of Management Information Systems, College of Commerce National Chengchi University and Airiti Press Inc, Vol. 18, No. 1, September (2012), pp. 57-69.

[3] H.Marghny Mohamed and I.Hussein Abul-kasim, Data Hiding by LSB Substitution using Gene Expression Programming,International Journal of Computer Applications (IJCA), Vol.45 - No. 14, 2012.

[4] Marghny Mohamed, Abeer Al-Mehdhar and Mohamed Bamatraf, SOM PAD: Novel Data Security Algorithm on Self Organizing Map, Computer Science and Information Technology (CS and IT),2012.

[5] H. Marghny Mohamed, M. Naziha AL-Aidroos and A. Mohamed Bamatraf, A Combined Image Steganography Technique Based on Edge Concept and Dynamic LSB, International Journal of Engineering Research and Technology, Vol.1 - Issue 8 (October - 2012)

[6] Marghny Mohamed, Fadwa Al-Afari and Mohamed Bamatraf, Data Hiding by LSB Substitution Using Genetic Optimal Key-Permutation ,International Arab Journal of e-Technology, Vol.2, No.1,January 2011.

[7] Ernatuti, R. Salim, Sulistyo, The applications of ELC numbers to golden cryptography, The Fifth International Conference on Information and Communication technology and Systems, 2009,p. 329334.

[8] M.Naziha Al-Aidroos, H. Marghny Mohamed, A. Mohamed Bamatraf , Hybrid Cryptographic Scheme for Data Communication , International Conference on Advanced Computer Theory and Engineering (ICACTE 2009).

[9] A.Rey and G. Sanchez, On the security of the golden cryptography, International Jounal of Network Security, Vol. 7, 2008, p. 448450.

[10] M.Naziha AL- Aidroos, H.Marghny Mohamed, and A.Mohamed Bamatraf, Data Hiding Technique Based on Dynamic LSB , Naif Arab University for Security Sciences..

[11] A.Nally, On the Hadamard product of golden matrices, Int. J. Contemp. Math Sciences, Vol. 2, 2007,p. 537 544.

[12] K.R.Sudha, A. Chandra Sekhar, Prasad Reddy PVGD, Cryptography protection of digital signals using some recurrence relations, Int. J.of Comp. Sci. and Network Security, Vol. 7, 2007, p. 203207.

[13] A.Stakhov, The golden matrices and a new kind of cryptography, Chaos, Solutions and Fractals, Vol. 32, 2007, p. 11381146.

[14] S.Kelly, S.Frankel, Using HMAC-SHA-256, HMAC-SHA-384, and HMAC-SHA-512 with IPsec, Internet Engineering Task Force (IETF), May 2007.

[15] A.Stakhov, Gazale formulas, a new class of the hyperbolic Fibonacci and Lucas functions, and the improved method of the golden cryptography, Moscow: Academy of Trinitarism, 2006, p. 11381146.

[16] S.William, Cryptography and Network Security Principles and Practices, Prentice Hall, fourth edition, 2005.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[17] K.Sayood, Introduction to Data Compression, 3rd Edition, Morgan Kaufmann, 2005.

[18] A.Jensen and A.la Cour-Harbo, Ripples in Mathematics The Discrete Wavelet Transform Springer-Verlag, 2001.

[19] Mallat Stephanie: A Wavelet Tour of Signal Processing. Academic Press, San Diego, 1998.

[20] S.Bruce,Applied Cryptography:Protocols, Algorithms, and Source Code in C. Wiley Computer Publishing, John Wiley & Sons, Inc. , second edition, 1995.

[21] CK.Chui, An Introduction to Wavelets. Academic Press, Boston; 1992.

[22] S.Vajda, Fibonacci and Lucas numbers, and the golden section. Theory and applications. Ellis Horwood limited; 1989.

[23] NN.Vorobyov, Fibonacci numbers. Moscow: Nauka; 1978 [in Russian].

[24] VE.Hoggat, Fibonacci and Lucas numbers. Palo Alto, CA: Houghton-Mifflin; 1969.

[25] C.Shannon, Communication theory of secrecy systems, Bell Systems Technical Journal, vol. 28, 1949, p. 656 - 715.

[26] N.Faller, An Adaptive System for Data Compression, In Record of the 7th Asilomar Conference on Circuits, Systems, and Computers. 1973, pp. 593-597.

[27] J. S.Vitter, Design and Analysis of Dynamic Huffman Codes, Journal of the ACM,vol. 34(4), 1987, pp. 825-845.

A Secure Attribute-based Model to Foster Collaboration in Healthcare Systems

Sara Najam
RITM Laboratory-ESTC
ENSEM-Hassan II University
Casablanca, Morocco
najam.sara@gmail.com

Hajar Mousannif
LISI Laboratory-FSSM
Cadi Ayyad University
Marrakesh, Morocco
mousannif@uca.ma

Mohamed Ouzzif
RITM Laboratory-ESTC
Hassan II University
Casablanca, Morocco
ouzzif@est-uh2c.ac.ma

***Abstract*—In today´ s rapidly-evolving globalized world, there is an undeniable trend towards establishing secure distributed collaborative work environments. In the prototypical example of health care institutions, critical research needs involve both effective collaboration modeling, and highly secured dynamic interactions insurance. In this paper, we introduce a new system design that enables both synchronous and asynchronous secure communication between different entities in a collaborative work environment. The proposed system provides a fine-grained attribute-based access control model to secure the collaboration in distributed systems, namely in Computer Supportive Cooperative Work (CSCW) systems. We opted for breast cancer diagnosis as a case study to apply our system design. Through a clear specification model of our system, we highlight the feasibility of achieving real-time and secure breast cancer diagnosis process, in which several medical organizations are engaged.**

Keywords— Healthcare; CSCW systems; ABAC; Collaboration; Security

I. Introduction

A lot of research has been recently driven by the goal of creating collaborative distributed systems which would allow groups of entities to cooperate together, work more productively, and share information in ways that have not been previously possible. The logic behind cooperation is that each participating entity is gaining more by cooperating, regardless of the extent, than if they operated alone or independently [1]. Distributed Computer Supportive Cooperative Work (CSCW) system is considered by many system designers and developers as the most convenient model for enabling effective collaborative environments and is widely used in many interdisciplinary application areas, including collaborative design and development, and conferencing [9].

CSCW as a design-oriented field has been particularly concerned, since its early days, with healthcare [3]. Healthcare organizations are an example of organizational systems that are composed of dozens of medical entities, and in which the proper functioning requires the alignment of several tasks performed by multiple departments together within a set of activities. However, any kind of usage of these distributed systems can raise many security issues and any redesign of care delivery should come with an eye on effective security management.

Role Based Access Control (RBAC) [2], which consists of predefined roles that have a set of privileges and to which subjects are assigned, has gained much popularity with respect to access control management within organizations, but it still suffers from many limitations in terms of flexibility [10].

In this paper, we propose a new system design that allows real-time and secure collaboration between different medical departments within healthcare organizations. The proposed collaborative system provides both synchronous and asynchronous secure interactions between entities, belonging to different organizations. As a proof of concept, we apply our design to the case of the most common invasive cancer in women worldwide which is breast cancer, and which has hit thousands of women since 2011 in UK alone [15] and other thousands of them in the US [19].

While all agree that designing such a system can be highly beneficial, addressing the following problem areas is a challenging task:

- What are the requirements to build a distributed system that can provide both synchronous and asynchronous collaboration for an effective breast cancer diagnosis?
- Why RBAC model is not enough to secure the medical systems based on CSCW model?
- How can we improve the RBAC model to have a fine-grained access control in CSCW system?
- How to make the access rule to an object flexible according to the user, object and environment attributes?

The remainder of this paper is organized as follows: In section 2, we review some related work in the area of distributed CSCW systems. Section 3 presents the requirements on which our proposed model is based. In section 4, we present our attribute-based model for CSCW systems. In section 5, we introduce our ABAC model specified by a breast cancer use case. Section 6 presents the implementation process of the collaborative application and section 7 concludes the paper and gives directions for future work.

II. Related work

CSCW research has long been interested in healthcare [4]. Many researchers have focused on the hospital setting. A recent research area has been devoted to understand collaboration in health care and to design technologies that support highly collaborative work of health care professionals [5]. On

another aspect, the world faces significant challenges related to security. These challenges have led researchers to develop models to secure distributed and collaborative systems.

Many access control (AC) models have been proposed to secure collaborative systems. These models are reviewed and compared in [14] . In addition to the basic model RBAC [2], the paper presents some other extended models including Team Based Access Control model (TMAC) [12] which was proposed as an approach to apply role-based access control in collaborative environments.

Other similar models of access control policies have been developed after TMAC occurrence [13]. For example, [15] [16] gives a description of OrBAC (Organization based AC) model which introduces the notion of organization where subjects are seen as roles, operations as activities, and objects as views. Each organization has its own policy, same for TMAC teams.

TMAC and OrBAC models add complexity in the design of access control policies and lack of dynamicity. Such policy models seem inadequate in dynamic environments. Therefore, some authors introduced administrative extensions (i.e. Ad-OrBAC [16]) or temporal models (e.g., Temporal RBAC(TRBAC), [17]) that can activate and deactivate some roles dynamically.

Smari et al. [13] presented an extended attribute-based access control model for collaborative management systems used for crisis management. The extended model is based on attributes associated with subjects and objects to address trust and privacy issues.

Tanvir [9] has developed a Role-based specification model for programming distributed CSCW systems. It provides dynamic security for such systems. Role activation, admission and validation are controlled through constraints. These constraints are defined and enforced within the role which makes the access control mechanism decentralized and dynamic.

Kulkarni and Tripathi [11] have proposed an extension of the NIST RBAC model [7] for addressing context-based access control requirements [11]. In addition to access control layer, the approach proposes a separate layer for context management. It supports personalized role permissions that are based on context information i.e. user location. The context-aware layer allows dynamic Role-Permission assignment to users. Role revocation is also supported, when conditions no longer satisfy the constraints.

Isabel et al. [6] share similarity with our approach in using attributes to enhance access control model in collaborative systems. In their work, context information is considered as user attributes. Constraints on user attribute values are used to assign users to the correct role, based on the values of different attributes. Therefore, users actions are dynamically determined based on their own attribute values and on the values of the attributes associated with the resources.

With respect to all related efforts mentioned above, very few researches have focused on tackling security issues in CSCW systems. Our work proposes a new access control model for building secure CSCW systems.

- We propose a new system design that provides efficient collaboration between engaged entities.
- We propose fine-grained access control based on ABAC to secure the collaboration.
- Our system design supports personalized and dynamic permissions for users to perform operations on objects.
- The request decision is based on the validation of several conditions that rely on attributes characteristics to be satisfied.

III. Security Requirements For Attribute-Aware CSCW Systems

RBAC has been widely used to express the access control constraints in organization systems. The management of access control in such systems is central. Roles creation tends to be based upon static organizational positions and users are statically assigned to roles by an administrator. Such model cant fit the requirements of distributed systems design. Our system design is based on the role-based model proposed by [9].Which is designed for programming distributed CSCW systems. Users are admitted to a role based on some constraints specified within the role. That makes the access control management dynamic and decentralized. We give more details about the RBAC model in the following section

A. Role Model in Distributed CSCW Systems

1) The Role Model Specification: The model is composed of a succession of activities involving a group of users. Each activity shows collaboration between several users; belonging to the same or different groups; to perform cooperative tasks. The group named Role is composed of users having common characteristics; these users have privileges to perform some actions upon specific resources. An activity is created and started by instantiating its template. The activity template defines several elements

- A set of roles that are engaged to do collaborative tasks.
- A set of operations associated with each role; the users who belong to the role can execute the methods provided by the operation. The operation can be a method executed on a shared object that is specified in the activity, or the creation of new nested activity.
- A set of object types that are accessed and created by the roles users through the operations.
- A set of constraints that specify for a user a set of conditions that should be satisfied in order to perform an operation or to admit his membership in a role.
- A set of child activity templates that are created within the activity. The child activity template can be instantiated when the operation that contains the template specification is invoked.

By creating the activity, the roles involved in the collaboration are specified in the activity template, as well as the object needed to be shared within it. The activities defined in the scope of their parent activity are automatically considered as nested activities, the objects created in a nested activity are accessible by the parent activity roles.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The users can be dynamically assigned to the roles in different ways. Role constraints are defined to control users membership as well as the role activation. These constraints are specified and enforced within their corresponding role.
The concept of meta-role called The OWNER is included in the proposed model. The user who instantiates the activity is automatically the owner of it. The owner role is responsible for managing all nested activities. It has the right to access the objects created or shared in the nested activities. The role is also the owner of the objects he creates and can grant this privilege to other roles.
In a distributed CSCW system, we cannot trust there is a single participant, or a role, or a domain to serve as a reference monitor to enforce all of the security policies. Instead, the model design specifies for each entity (object, role, and activity) a role that can be trusted to correctly enforce its management functions.

2) Role Constraints in CSCW model: In a distributed CSCW system, access control policy is required to securely manage activities. It provides an access control to roles and objects in the activities. Security requirements in the CSCW system are naturally dynamic. They depend on the execution history of the collaborative tasks and temporal conditions. Security policy is presented as con-straints that should be satisfied by the user to access a role or perform an action. These constraints can be based on users history membership in roles. The user can be admitted to the current role if he was member in prerequisite previous roles and denied to access the role if he was a member in conflicting ones. The constraints can be also based on history of past actions performed by a user, and role membership cardinality.
For role security management, three types of role constrains are defined: role admission, validation and activation constraints. Operation preconditions are also requested to manage the operation execution within the role.
The following paragraph gives more details about the constraints and the operation precondition, as well as conditions types.

- **Admission constraints:** It specifies the conditions that must be satisfied for a user to be admitted in the role.
- **Validation constraints:** Once the user is admitted into the role, the admission constraints no longer hold. Role-revocation condition is needed to verify if the current user s membership in the role is still valid. The validation constraint is used to check the validity of user s membership.
- **Activation constraints:** specify the conditions that should be satisfied to keep the role active. The condition is temporal when the service provided by the role is available in a specific time or the access to an object by the role members is restricted to limited time. The activation condition can also be based on cardinality; it specifies the lowest number of members that must be admitted in the role before any execution of the role operations.
- **Operation preconditions:** are used to enforce the access control policy and allow or deny a user to perform an operation.

B. Why RBAC Model is not sufficient for Distributed CSCW Systems

Despite its popularity and wide use in distributed systems, the RBAC model has many limitations in term of flexibility. In this section, we highlight some these limitations.
In RBAC model, access to an object is granted to locally defined roles that the subject is member of. That means that all members of a defined Role have the same permission to access an object. This approach is not enough when the access to a specific object is selective and limited only to some members of the role according to some features. E.g. in medical domain when a complicated case is presented to the hospital, only Doctor who have more than 10 years of experience can intervene in such case.
RBAC makes a decision based on the subjects association with a role. RBAC does not easily support multi-factor decisions [10]. A decision based on subjects membership in a role is essential but not enough to have a fine-grained access control. Taking an access decision over an object requires a validation of many factors: the subject constraints, the object constraints and also the context and the environment in which the access request occurs.
RBAC role assignments tend to be based upon static structures predefined by an administrator. This may induce challenges in certain RBAC architectures where dynamic access control decisions are required. A role creation or removal can be hard to manage by the administrator after the model implementation. Trying to implement the role-based access control decisions would require the creation of numerous roles that are ad hoc and limited in membership, leading to what is often termed role explosion [10].
As mentioned in section III.A.1, the owner of an activity or object has the right to grant the privilege to other role using the predicate CHANGE-OWNER. This privilege should be granted according to some constraints. The role to which the privilege will be granted should have some attributes that satisfy the conditions specified in the constraints. After doing that, the owner has no longer the right to retrieve this grant or to revoke the ownership from the new owner.

IV. ATTRIBUTE BASED MODEL FOR CSCW SYSTEMS

In order to address the aforementioned limitations, we propose an extension of the RBAC model based on attributes. We use Attribute-based access control (ABAC) model.

A. ABAC Concepts

According to NIST [10], the high-level definition of ABAC:

An access control method where subject requests to perform operations on objects are granted or denied based on assigned attributes of the subject, assigned attributes of the object, environment conditions, and a set of policies that are specified in terms of those attributes and conditions.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Subject is a user to whom a set of attributes are assigned. He may be member of a role and through this role he can perform an operation or access to an object. **The Object** is the resource on which the access control is performed (e.g. record, file, etc.). It can also be the creation of an activity or the execution of an operation. **The policy** is the representation of rules that specify the decision either granting or denying the access to subject. The decision is based on the values of the attributes of the subject, object, and possibly environment conditions.

Environment conditions are the operational or situational context in which access requests occur. Environment characteristics are independent of subject or object. This and may include the current time, day of the week, location of a user, or the current threat level [9]. In the literature, an approach has been proposed to handle the context concept in the RBAC model, especially the context of location [10]. In our model, we consider the location context as an attribute in the environment conditions, in addition to the time attribute as well as the conditions of the executing operational environment.

Next section describes how the attribute-based access control is modeled in our proposed model.

B. Proposed ABAC Model for Distributed CSCW Systems

In our model, decision is based on the verification of the Subject attributes. This means that the permission to access an object differs from one member to another within the same role. Hence, we added an attribute layer to the role, in which the attributes of the members assigned to this role will be defined. The subjects attributes are managed by a role manager. We also added an attribute layer in the objects specification, so the objects attributes are defined within the same object and managed by an object manager.

The model presents different constraints that make the access control more restricted than RBAC Fig. 1. The admission constraints can be used to control the users admission to a specific role; e.g. in breast cancer diagnosis activity, when gynecologists discuss and give their decision about the case. Only Intern students have the right to assist the activity. The attributes on which admission constraints to gynecologist role will be based is status.

The validation constraint is used to check the membership of a user when he requests the execution of an operation. The validation constraints specify whether the user should be revoked from the role or not, according to his attributes. E.g. within a diagnostic activity, the validation constraints specify that the users membership in the gynecologist is revoked if she becomes member of patient role. A gynecologist cannot diagnose herself. As we mentioned in section III.A.2, the activation constraints verify the conditions that must be satisfied to keep the role active. In our model, the conditions will be based on the subject attributes as well as the context-based condition, i.e. subject location or time.

The subjects request to perform an operation within a role requires the validation of different constraints through different layers. The attribute-based access control mechanism examines the subjects attributes through admission and validation constraints. It also examines the objects attributes. The Access control mechanism then determines what operations the subject may perform to create an activity or access an object. The decision is specified in the operation precondition. In the literature, an approach has been proposed to manage the context-based conditions by integrating a context-based management layer in the RBAC model. Those context-based conditions can be specified in the constraints and precondition mechanism to define the request decision.

V. Model specification

CSCW as a field has been concerned since its early days with healthcare and is widely used in collaborative systems where different entities collaborate together in order to make decisions about a patient ' s case.

A. Breast Cancer Scenario

We illustrate our proposed model using a breast cancer case study. Diagnosis procedures for breast cancer require the collaboration between several doctors from different specialty departments.

- When a patient suspects having a breast cancer, she goes to hospital to determine the cause of the symptoms she has.
- A general practitioner (GP) performs a careful physical exam that includes a personal and family medical history. In addition, an examination may be done which includes palpation - carefully feeling the lump and the tissue around it - its size, its texture, and whether it moves easily. Benign lumps often feel different from cancerous ones.
- In addition to the physical examination by the GP, the patient is sent to radiology department for other imaging tests. Imaging tests may include one or more tests like mammography. GP recommend a diagnosis mammogram from the Radiologist to further evaluate the case.
- Based on these exams, the GP may decide that no further tests are needed and no treatment is necessary. In such cases, he/she may ask to check the patient regularly to watch for any changes Otherwise, in cases of suspected cancer, the GP transfers the patient to a specialist physician (Gynecologist) and shares the patient record, imaging tests and all examination reports with him/her.
- The gynecologist removes fluid or tissue from the breast to be sent to the Anatomo-pathology laboratory to look for cancer cells. The procedure is called a biopsy. The biopsy can be done using a needle to get a piece of the area of con-cern, or it can be done with surgery which requires the intervention of a surgeon.
- A biopsy removes tissue or cells from the body for examination under a microscope. It is the only way to determine for sure if cancer or other abnormal cells are present.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

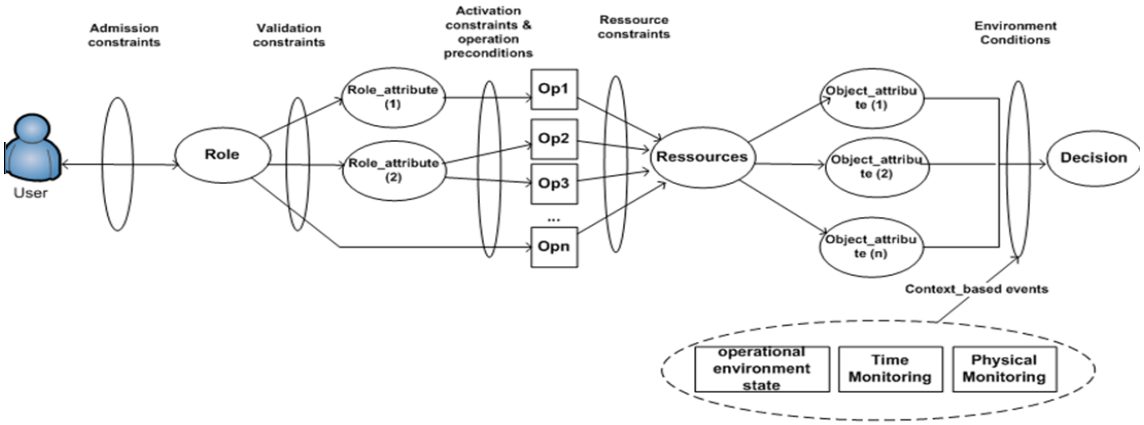


Fig. 1. Attribute-based model for CSCW systems.

- The Anapath lab studies the cells and gives the gynecologist the exact and final report. It determines whether the cells are cancerous, the type of cells involved in the breast cancer, the cancer grade, etc. these results will be used to guide therapy.
- Once the gynecologist diagnosed the breast cancer, he/she prescribes other tests to the patient to determine the stage of the cancer. The staging consists of a breast MRI, Mammogram for the other breast to look for signs of cancer, an abdominal and pelvic echography, a bone scan and blood tests.
- The radiologist sends the staging results to the gynecologist. The gynecologist determines the stage of the cancer ac-cording to the T.M.N strategy.
- The gynecologist creates a collaborative session in which she will discuss the patient case with other gynecologists and have their opinion about it.
- Each engaged gynecologist give her/his experience and decision about the case. And according to all decisions given by the doctors, the responsible gynecologist gives her final diagnosis and decides whether the patient needs surgery, chemotherapy or even other therapies.
- A final collaborative session is created to gather the gynecologist, the surgeon and chemo-therapist in order to discuss the patient case and start the therapy.

B. Model specification

Our system model can be specified using a hierarchical structure of activities in which physicians are specified as subjects and grouped into roles according to their specialty. Within each role, members are associated to their attributes by which they will be selected to perform operation.
In our model Fig. 2, the breast cancer diagnosis process can be specified as a set of activities templates. In each activity, doctors from different specialties cooperate to do a part of the diagnosis. The main activity template *Diagnosis* includes all other activities; it presents the global concept of diagnosis in the hospital. Diagnosis contains two roles patient and physician, and one shared object called *patient_record*. Each member of the physican role can instantiate the nested *Intial_Diagnosis* activity when he wants to examine a patient. A partial specification of Diagnosis activity is presented as follow (Alg.1)

In case of patient examination, the *initial_diagnosi* activity

Algorithm 1

ACTIVITY_TEMPLATE Diagnosis(**OWNER** physician,
OBJECTS PatientRecord p_record
ASSIGNED_ROLES physician, Patient)
{
Role physician { ...}
Role patient { ...}

template is instantiated by the physican member. Mostly the physician member is the GP who is responsible for the patient diagnosis. The initial_diagnosis activity is defined with three roles: *patient*, *general_practitioner* and *radiologist*. The patient, the general practitioner responsible for the patient diagnosis and an available radiologist are assigned to *patient*, *general_Practitioner* and *Radiologist* Roles respectively. The members assignment to the nested activity roles can be performed in two ways: the members of a role in the parent activity are statically assigned to a role in child activity using **Reflect** tag ; the patient role in Diagnosis activity is statically assigned to patient role in *Initial_diagnosis*. The other way to assign members to the role is by using **AssignedRole** tag in the activity specification template. The role to which some members should be assigned is specified within the tag.
In our example, the *general_practitioner* role has four operations; *startDiagnosis* allows to the GP to see the symptoms and examines the patient, the operation action in this case the invocation of *writeSymptoms* method on *patient_record* object. If the presence of a breast cancer is suspected, the GP executes the second operation

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

MammoRequest asking the radiologist to perform a mamo test for the patient. To do so, the operations action will be the creation of a new request object. The third available operation for the GP role is *DoEchography*. The GP executes it when the Mammo test shows doubts about the cancer to have more accurate results. The last operation *TransferPatient* is executed when the GP is more confident that the patient has a breast cancer. The GP performs the operation ' s action which consists of creating a new activity Specialist_Diagnosis, in which the patient will be transferred to a gynecologist specialist who will take in charge the patient for more tests and eventual therapy. By creating the activity, the roles involved in the collaboration are specified in the activity template, as well as the object needed to be shared in this activity e.g., patient_record, Mammo and Echography reports (Alg.2)

In RBAC model, the activities defined in the scope of their parent activity are automatically nested and are also called child activities. The objects created in a nested activity are available to the parent activity ' s roles. In our proposed model, the activity specified in the scope of the parent activity can be a separate activity that will be executed independently from the parent activity. Which is the case of *Specialist_Diagnosis* activity that is defined within *Initial_Diagnosis* activity but both of the activities are executed successively with separate roles.

Algorithm 2

```
ACTIVITY_TEMPLATE Initial_Diagnosis(
OWNER General_Practitioner,
OBJECTS PatientRecord p_record,
ASSIGNED_ROLES General_Practitioner,
Radiologist, Patient)
{
...
Role General_Practitioner {
Operation StartDiagnosis {
PRECONDITION #StartDiagnostic.start
(invoker=thisUser) = 0
ACTION {
Grant p_record readFile;
Grant p_record writeSymptoms;
}
}
...
}
...
}
```

An instance of the *Specialist_Diagnosis* template is created for each patient transferred to Gynecology department and a gynecologist member is assigned to his matching role. An *anatomo_pathologist* is also assigned to its role in order to help the gynecologist doing the diagnosis. The patient is also assigned within the activity to be able to access her record. Within each *Specialist_Diagnosis* activity instance, the gynecologist who is the owner of the instance can create some nested activities to collaborate with other entities in order to determine the stage of the cancer as well as the therapy. The gynecologist role is responsible for managing all nested activities, it has the right to access the objects created or shared in the nested activities. The specification of the owner role is defined by **OWNER** tag in the nested activity template specification(Alg.3).

Algorithm 3

```
ROLE General_Practitioner {
OPERATION Patient_Transfer {
PRECONDITION #(DoEchography.finish)> 0
ACTION {
OPERATION session = new ACTIVITY Special-
ist_Diagnosis
PassedObject p_record, mammo, eco
MemberAssignment PatientPatient= ThisActivity.Patient ;
...
}
}
}
```

The first nested activity called *Extended_Assessment* is created for more tests to determine the stage of the cancer, the gynecologist role initiates the activity and cooperates with the radiologist role and the patient role is added for the patient to be able to see the shared test results. After knowing the cancer stage, the gynecologist role creates a second nested activity *Collaborative_Diagnosis* in which more gynecologist members will be assigned to the role in order to collaborate with the owner and discuss the patients case to have a final decision.

Within the *Collaborative_Diagnosis* activity instance, there are two types of role coordination. Coordination between Gynecologists and Radiologists members is called **interrole** coordination. The second one is called **intrarole** coordination and is required when many members of the same role need to coordinate among themselves. This is the case of gynecologist role in *Collaborative_Diagnosis*. The intrarole coordination can be based on independent, cooperative or ad-hoc modes for the role ' s task execution by the members. In our example, the coordination mode is cooperative because the gynecologist who is assigned to manager role will cooperate with the other gynecologists to have a final decision. After the cooperation, the gynecologist role creates the T.M.N classification object that contains all description about the patient ' s disease and the guide of therapy as well. The gynecologist creates a final activity instance named therapy. It contains a set of roles that will take care of the patient Therapy.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

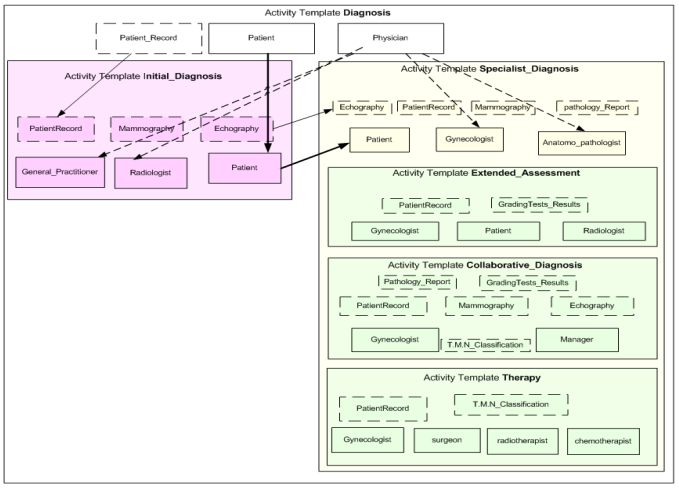


Fig. 2. Activities Hierarchy of Breast Cancer Diagnosis Process

C. subject Attribute

Three types of role constrains are defined: role admission, validation and activation constraints, and operation preconditions to manage the operation execution within the role. These constraints are defined in the specification model using three kinds of conditions: Role Membership Functions that check and manage the user membership in a specific role, Event Based predicates, and temporal conditions. The different constraints can use any type of the listed conditions.
The following paragraph gives more details about the constraints and the operation precondition, as well as conditions types. We illustrate different aspects of security requirements using *Collaborative_Diagnosis* activity as example.

1) Attribute-Based Admission Constraints: In Gynecologist role, the admission constraint uses the Role Membership Functions (RMF) type of condition. RMF is a set of functions used to verify the user membership in the role. *member(thisUser, Parentactivity.Physician)* is a Boolean function that checks the user membership and returns true if the current user is a member in the physician role in the parent activity (which is Diagnosis activity). *members (ThisRole)* returns the list of all current members of the role.We can count the members list by using the operator # on the previous function *#members (ThisRole)*. This type of condition is not limited to the admission constraint, it can be used in activation and validation constraints. As we mentioned earlier, the access control requirements are based on the role concept which is essential but not enough in a distributed system, especially the medical domain. We propose an Attribute-based access control to be used in the role constraints and the operation preconditions. We can extend the Role Membership Functions to be Attribute Membership Function. The extension of member function will be defined as follow: *Attributes(ThisRole)* that lists all the current role ' s attributes. The operator # gives the number of attributes that the current role has. *Attributes(Role, ThisUser)* lists the Role ' s attributes with their corresponding values. The function *Attribute (ThisUser, attribute_name)* returns the value of attribute_name that user has. The function is used to take the decision either granting or denying access to the user . E.g, Attribute (ThisUser, Experience)>10 checks if the current user ' s experience is more than 10 years. We can specify (Specialty, gender, experience, location) as attributes on which the access control to activities and objects can be based. The other function *has_attr(ThisUser.attribute, attribute_value)* checks if the value of the user ' s attribute matches the one specified in the function, e.g *has_attr(ThisUser.Speciality, Gynechology)* returns true if the current user is gynechologist and false if he is not . These predicates can be used in operation preconditions and role constraints(Alg.4).

Algorithm 4

ROLE Gynecologist {
ADMISSION_CONSTRAINTS member(thisUser, ParentActivity.Physician)
∧ has_attr(ThisUser.Speciality,Gynechology)
∧ Attribute (ThisUser, Experience)> 10

2) Attribute-Based Validation Constraints: The validation constraint is used to check the validity of user ' s membership. Validation constraint of the Gynecologist role specifies that the user ' s membership is revoked if he becomes member of patient role. The validation constraints mostly contain Role/Attribute Membership Function condition (Alg.5)

Algorithm 5

ROLE Gynecologist {
VALIDATION_CONSTRAINTS
!member(ThisUser,Patient)
Operation ReadSymptoms {
ACTION {
Grant p_record readFile;
Grant eco readResults;
Grant mammo readResult;
}
}
...
}

3) Attribute-Based Activation Constraints: The condition is usually temporal when the service provided by the role is available in a specific time or the access to an object by the role members is restricted to limited time e.g. *time>DATE(May, 14, 2014, 9:00)* ∧ *time<DATE(May, 14, 2014, 11:00)*. The second condition is based on cardinality. It specifies the lowest number of members that must be admitted in the role before any operation is executed(Alg.6).

D. Resource and Environment Attributes

1) Attribute-Based Resource Constraints: In case the user sends a request to access an object, the resource attributes checking become also essential for a request decision to be

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Algorithm 6

ROLE Gynecologist {
...
ACTIVATION_CONSTRAINTS #members(ThisRole)>0
OPERATION readSymptoms {
...
}
}

taken. We propose functions that allow assigning or getting attributes from an object. In Collaborative_Diagnosis activity, a *TMN_classification* object is created by the gynechologist to give the final decision about the patient ' s case. The function *set_attr(ThisObject, attr_name, attr_value)* assigns to the object the specified value and *get_attr(ThisObject, attr_name)* returns the value of the object ' s attribute (Alg.7).

Algorithm 7

OPERATION FinalDecision {
ACTION {
Tmn = new OBJECT TMN_Classification;
set_attr(ThisObject, name, Alice_tmn);
Grant tmn setFinalDecision;
}

2) Context Management Layer: Some researchers have proposed a separate context management layer [11]. Through this layer, we can manage the ambient context by sensing various kinds of conditions in the environment i.e. user location, etc. in addition to ambient context, we can also take into consideration internal context information like time execution of the application. The context conditions are expressed by predicates and evaluated by Role/ Object managers in time of request validation.

VI. Implementation

In this section, we present the implementation process of our proposed system design. The process allows setting up the collaborative application; it takes as input the specification model and provides at the end a visible and clear view of the model architecture. The schema below Fig. 3 describes the main three steps of the process, the creation of XML-based schema of the system, generating Java classes and the creation of the collaborative application based on the database repository.

The first step consists of creating an initial xml schema of the application. The structure of the schema is based on the given specification model Fig. 4. The main elements of the model specification section III.A are specified in the schema as elements of xml schema definition (XSD) [20]. The liaison and cardinalities of these objects are also specified in the xml schema.

The second step of the implementation uses Java Architecture

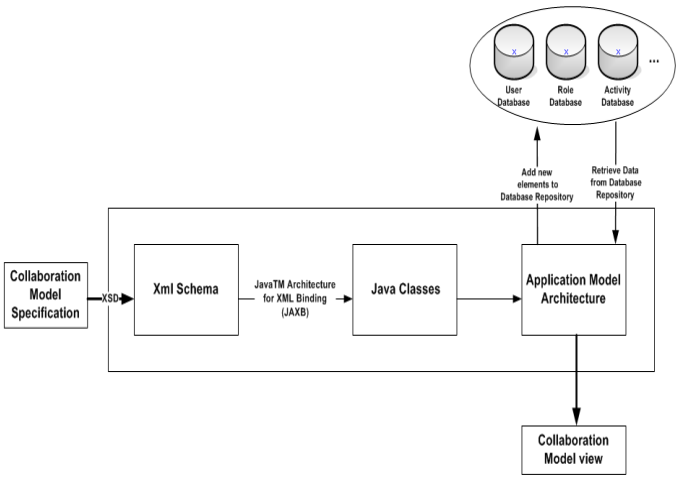


Fig. 3. The implementation process of the collaborative application

for XML Binding (JAXB) [21] to automatically generate java classes from Xml schema. JAXB is particularly useful when the specification is complex and changing [21]. Each element in the xml representation will be converted to a corresponding java class. The types of xml schema elements are mapped to Java data types and the liaison between the elements is conserved as well for the corresponding classes. This step of the implementation process is very important as it makes the creation of the application java classes transparent and dynamic. We can simply generate a new version of the classes every time the specification changes. The Java classes form the application skeleton. The last step of the implementation consists of retrieving data from databased and instantiating the classes to develop the application code. We firstly developed an editor that allows to the user a visibility on different elements of the collaboration as well as the model structure. It also allows to the application user the possibility to add or remove an element from the tree view Fig. 5.

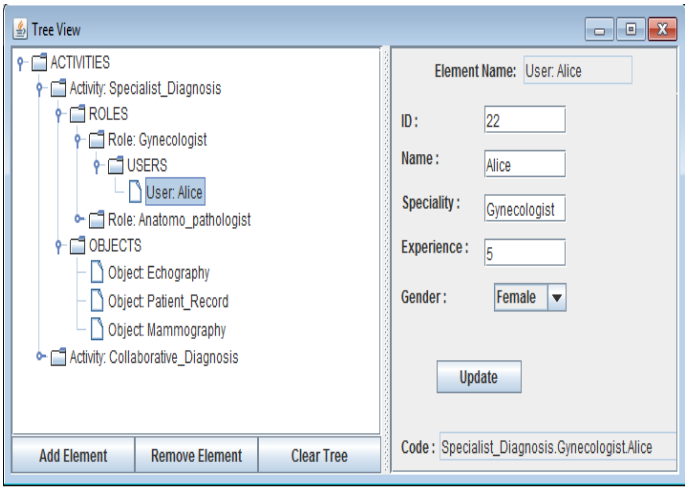


Fig. 5. The TreeView of the collaborative application

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

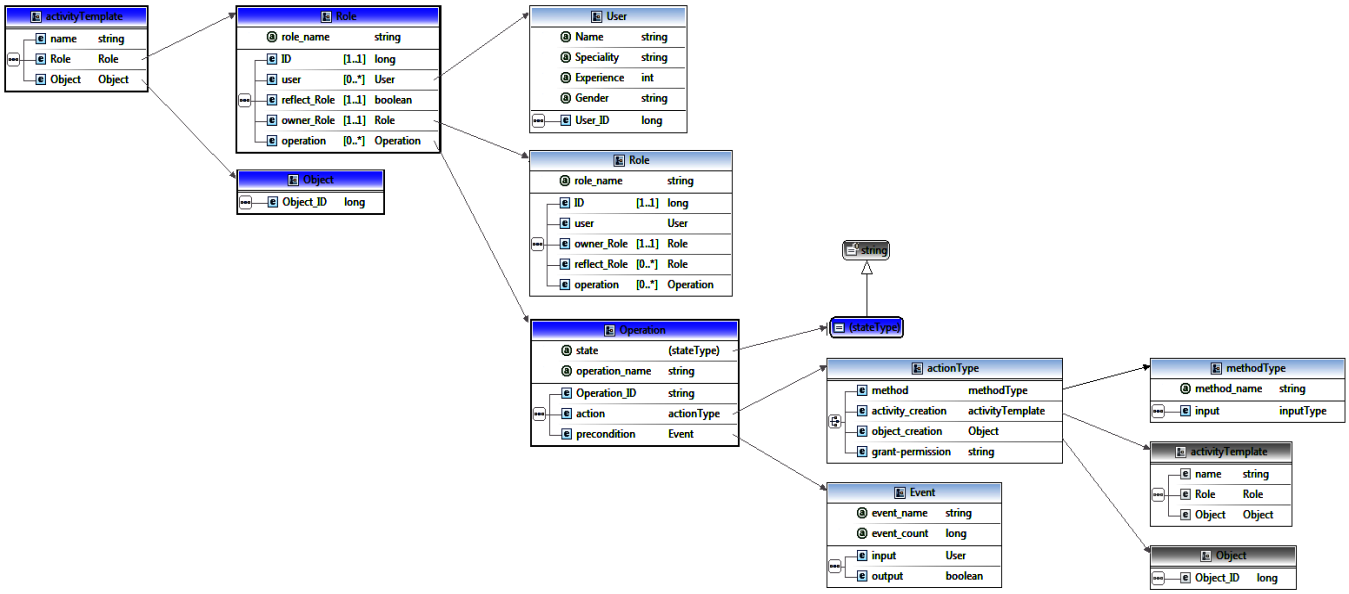


Fig. 4. XML schema of the collaborative application

VII. CONSLUSION

In this paper, we proposed a new system design that provides a fine-grained access control model for constructing secure distributed CSCW systems. The system design allows different entities to collaborate more effectively and securely. We highlighted the limitations of RBAC model of such distributed system. We also demonstrated the need of an attrib-ute-based access control for more flexible access control management. The system design supports collaboration by providing a succession of activities. These latters contain different groups of users that can have synchronous and asynchronous interactions within the activity to perform common tasks. We also presented the key elements of our model specification through a breast cancer diagnosis scenario. Future work will consist in incorporating activity ordering in the proposed system design and establish a framework for programming secure attribute-based distributed CSCW systems.

REFERENCES

[1] H. Mousannif, I. Khalil, and G. Kotsis, "The cloud is not there, we are the cloud!, International Journal of Web and Grid Services 9(1), 117, 2013.

[2] R. Sandhu, E. Coyne, H. Feinstein,and C. Youman, "Role-based access control models, IEEE Comput., 29, (2),, Feb 1996.

[3] G. Fitzpatrick, and G. Ellingsen, "A Review of 25 Years of CSCW Research in Healthcare: Contributions, Challenges and Future Agendas, Computer Supported Cooperative Work (CSCW)", Journal of Collaborative Computing and work practices, Springer,2012.

[4] C. Heath, and P. Luff, "Documents and Professional Practice: 'bad' organisational reasons for 'good' clinical records", In Proc. of Computer-Supported Cooperative Work (CSCW). Boston, MA USA. pp: 354-363. ,1996

[5] M. Reddy, J. Bardram, and P. Gorman, "CSCW Research in Healthcare: Past, Present, and Future ", In Proc. of Computer-Supported Cooperative Work (CSCW) ,2010

[6] F. Isabel, G. Cruz, R. jomemo, B. Lin, and M. Orsini "A Constraint and Attribute Based Security Framework for Dynamic Role Assignment in Collaborative Environments. Collaborative Computing: Networking, Applications and Worksharing", LNIC Sciences, Social Informatics and Telecommunications Engineering Volume 10, pp 322-339 , 2009

[7] D.F. Ferraiolo, R. Sandhu, S. Gavrila, D.R. Kuhn, and R. Chandramouli, "Proposed NIST standard for Role-based Access Control", ACM Transactions on Information and System Security (TISSEC), 4(3):224274,2001

[8] J. Abraham, and M.C. Reddy, "Re-Coordinating Activities: An Investigation of Articulation Work In Patient Transfers", CSCW'13,2013

[9] T. Ahmed, and A.R Tripathi, "Specification and Verification of Security Requirements in a Programming Model for Decentralized CSCW Systems", ACM Transactions on Information and System Security (TISSEC), 10(2),2007

[10] V. C. Hu, D. Ferraiolo, R. Kuhn, A. Schnitzer, K. Sandlin, R.Miller, and K.Scarfone. "Guide to Attribute Based Access Control (ABAC) Definition and Consideration", NIST Special Publication 800-162 Natl. Inst. Stand. Technol. Spec. Publ. 800-162, 45, Jan 2014

[11] D. Kulkarni, and A. Tripathi, "A. Context-aware role-based access control in pervasive computing systems", In Proceedings of the 13th ACM Symposium on Access Control Models and Technologies (SACMAT08), pages 113122, Estes Park, CO, USA, June 2008

[12] R.K. Thomas, "Teambased Access Control (TMAC): A Primitive for Applying RoleBased Access Controls in Collaborative Environments", Proceedings of the Second ACM Workshop on Role-Based Access Control, Fairfax, Virginia, USA, pp.13-19, November 06-07, 1997.

[13] W.W Smari, P. Clemente, and J.F. Laland, "An extended attribute based access control model with trust and privacy: Application to a collaborative crisis management system", Future Generation Computer Systems Volume 31, Pages 147168, Feb 2014
[14] W. Tolone, G., Ahn, and T. Pai, "Access control in collaborative systems", ACM Computing Survey. 37, 1 , March 2005.
[15] F. Cuppens, and A. Mige, "Modelling Contexts in the OrBAC Model", The 19th Annual Computer Security Applications Conference, Las Vegas, Nevada, USA. 2003
[16] F. Cuppens, and A. Mige, "Administration Model for Or-BAC,", International Journal of Computer Systems Science and Engineering (CSSE).2004.
[17] E. Bertino, P.A. Bonatti, and E. Ferrari, "TRBAC: A Temporal Role-Based Access Control Model,", ACM Transactions on Information and System Security (TISSEC), 4(3), pp. 191-233, 2001.
[18] http://www.cancerresearchuk.org/cancer-info/cancerstats/types/breast/
[19] http://www.breastcancer.org/symptoms/understand_bc/statistics
[20] http://fr.wikipedia.org/wiki/XML_Schema
[21] http://en.wikipedia.org/wiki/Java_Architecture_for_XML_Binding

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

A two-stage architecture for stock price forecasting by combining SOM and fuzzy-SVM

Duc-Hien Nguyen, Manh-Thanh Le
Hue University
Hue, VietNam

***Abstract*— This paper proposed a model to predict the stock price based on combining Self-Organizing Map (SOM) and fuzzy – Support Vector Machines (f-SVM). Extraction of fuzzy rules from raw data based on the combining of statistical machine learning models is the base of this proposed approach. In the proposed model, SOM is used as a clustering algorithm to partition the whole input space into several disjoint regions. For each partition, a set of fuzzy rules is extracted based on a f-SVM combining model. Then fuzzy rules sets are used to predict the test data using fuzzy inference algorithms. The performance of the proposed approach is compared with other models using four data sets.**

Keywords- Fuzzy rules; Support vector machine - SVM; Self-Organizing Map - SOM; Stock price forecasting; Data-driven model

I. INTRODUCTION

Nowadays, time series forecasting, especially predicting the stock market has attracted a lot of interest from many scientists. For the ultimate objective of increasing the accuracy of predicting results, many researchers have made contributions to conducting and improving various models and solutions. The current stock market prediction is approached in two methods, either stock price prediction or the trend of stock price past n-days.

Today, the application of data mining and statistical machine learning techniques are two common approaches used to predict stock market movements. Many researches in [7], [8], [14], [16], [17] proposed applications of Artificial Neutral Network, Support Vector Machine - SVM, Hidden Markov Model - HMM in stock market prediction. In order to make more effective and accurate predictions, various combined models with different forecasting methods [4], [9], [11] are researched and proposed by researchers. A model based on the Fuzzy model combined with Support Vector Machine is a new trend of research, called data-driven model [5], [6], [10], whose purpose is to extract fuzzy rules from Support Vector Machine as basic functions for fuzzy system. One of the challenges for the data-driven model is automatic learning from data whose size is large but representativeness is limited. In addition, avoidance of the explosion in the number of fuzzy-rules is also a problem which needs solving.

In order to resolve the large sizes of the data set problem in data-driven model, combination of a data clustering algorithm such as k-Means, SOM,…is a new approach which used to divide the large sizes of the data in to several smaller sizes of data set [4], [11]. The main purpose of this study is to deal with the large size of the data, minimize the quantity, simplify fuzzy rules from the data; we propose a model combining SOM and SVM for fuzzy rule extraction in stock price prediction. The fuzzy rules set in small amounts will create favourable conditions for human experts to understand, dissect, evaluate, and optimize to improve the efficiency of fuzzy rules-based inference system.

The rest of this paper is organized as follows. Section 2 briefly describes the theory to support vector machines, fuzzy model and the relationships between the two models; then introduces the f-SVM method for extraction of fuzzy rules from SVMs. Section 3 presents SOM which has been widely used in data clustering. Section 4 introduces the proposed model which can produce fewer fuzzy rules based on the combination between SOM and f-SVM to predict stock market. The results obtained from the proposed model are demonstrated in comparison to other models, which will be presented in section 5. In section 6, we present the conclusion and future work.

II. FUZZY RULE EXTRACTION METHOD FROM SUPPORT VECTOR MACHINES – F-SVM ALGORITHM

Support vector machine (SVM), which is proposed by Vapnik, is a new machine learning method based on the Statistical Learning Theory and is a useful technique for data classification [2]. SVM has been recently introduced as a technique for solving regression estimation problems [5], [8], [11], and has also been used in finding fuzzy rules from numerical [5], [6], [10]. In the regression estimation task, the basic theory of SVM [2] can be briefly presented as follows:

Given a set of training data $\{(x_1, y_1), \dots, (x_l, y_l)\} \subset X \times \mathcal{R}$, where X denotes the space of input patterns. The goal of ε Support vector regression is to find a function f(x) that has at most ε deviation from the actually obtained targets y_i for all the training data, and at the same time is as flat as possible. That is, the errors would be ignored as long as they are less than ε, but any deviation bigger than this would not be accepted.

$$f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) K(x_i, x) + b \qquad (1)$$

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Subject to

$$\sum_{i-1}^{l}(\alpha_i - \alpha_i^*) = 0, \text{and } C \geq \alpha_i, \alpha_i^* \geq 0, \forall i, \quad (2)$$

Where, the constant C which determines the trade-off of error margin between the flatness of $f(x)$ and the amount of deviation in excess of ε that is tolerated; α_i, α_i^* are Lagrange multipliers; and $K(x_i, x)$ is a Kernel function defined as

$$K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle \quad (3)$$

where $\Phi(x_i)$ is a nonlinear function mapping.

The input points x_i with $(\alpha_i - \alpha_i^*) \neq 0$ are called support vectors (SVs).

On the other hand, fuzzy rule-base that generally consists of set of IF-THEN rules is the core of the fuzzy inference [5]. Suppose there are M fuzzy rules, it can be expressed as follows:

$$R_j: \text{IF } x_1 \text{ is } A_1^j \text{ and } x_2 \text{ is } A_2^j \text{ and } \dots \text{ and } x_n \text{ is } A_n^j \text{ THEN y is } B^j, \text{ for } j = 1..M \quad (4)$$

where $x_i (i = 1,2, \dots n)$ are the input variables; y is the output variable of the fuzzy system; and A_i^j and B^j are linguistic terms characterized by fuzzy membership functions $\mu_{A_i^j}(x_i)$ and $\mu_{B^j}(y)$, respectively.

The inference process is shown as below: 1) membership values activation. The membership values of input variables are computed as t-norm operator

$$\prod_{i=1}^{n} \mu_{A_i^j}(x_i) \quad (5)$$

2) the final output can be computed as

$$f(x) = \frac{\sum_{j=1}^{M} \bar{z}^j \left(\prod_{i=1}^{n} \mu_{A_i^j}(x_i) \right)}{\sum_{j=1}^{M} \prod_{i=1}^{n} \mu_{A_i^j}(x_i)} \quad (6)$$

where $\bar{z}^j$ is the output value when the membership function $\mu_{B^j}(y)$ achieves its maximum value.

In order to let equation (1) and (6) be equivalent, at first we have to let the kernel functions in (1) and the membership functions in (6) be equal. The Gaussian membership functions can be chosen as the kernel functions since the Mercer condition [15] should be satisfied. Besides, the bias term b of the expression (1) should be 0.

While the Gaussian functions are chosen as the kernel functions and membership functions, and the number of rules - M equal to the number of support vectors - l, then (1) and (6) become:

$$f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) \exp\left(-\frac{1}{2}\left(\frac{x_i - x}{\sigma_i} \right)^2 \right) \quad (7)$$

and

$$f(x) = \frac{\sum_{j=1}^{l} \bar{z}^j \exp\left(-\frac{1}{2}\left(\frac{x_j - x}{\sigma_j} \right)^2 \right)}{\sum_{j=1}^{l} \exp\left(-\frac{1}{2}\left(\frac{x_j - x}{\sigma_j} \right)^2 \right)} \quad (8)$$

The inference of fuzzy systems can be modified as [3]

$$f(x) = \sum_{j=1}^{l} \bar{z}^j \exp\left(-\frac{1}{2}\left(\frac{x_j - x}{\sigma_j} \right)^2 \right) \quad (9)$$

and the center of Gaussian membership functions are selected as

$$\bar{z}^j = (\alpha_i - \alpha_i^*) \quad (10)$$

Then, the output of fuzzy system (6) is equal to the output of SVM (1). However, the equivalence has some shortcomings: 1) the modified fuzzy model removes the normalization process; therefore, the modified fuzzy model sacrifices the generalization. 2) the interpretability cannot be provided during the modification.

An alternative approach is to set the kernel function of SVMs as

$$K(x_i, x) = \frac{\exp\left(-\frac{1}{2}\left(\frac{x_i - x}{\sigma_i} \right)^2 \right)}{\sum_{i=1}^{l} \exp\left(-\frac{1}{2}\left(\frac{x_i - x}{\sigma_i} \right)^2 \right)} \quad (11)$$

Consequently, the output of SVMs becomes

$$f(x) = \frac{\sum_{i=1}^{l}(\alpha_i - \alpha_i^*) \exp\left(-\frac{1}{2}\left(\frac{x_i - x}{\sigma_i} \right)^2 \right)}{\sum_{i=1}^{l} \exp\left(-\frac{1}{2}\left(\frac{x_i - x}{\sigma_i} \right)^2 \right)} \quad (12)$$

We only have to set the centre of membership functions to $(\alpha_i - \alpha_i^*)$, then we can assure the output fuzzy systems (12) and the output of the SVMs (7) are equal. Notably, the expression (11) can only be achieved when the number of support vectors, l, is known previously.

From the analysis of the similarity of SVMs and fuzzy systems above, we propose F-SVM algorithm in Figure 1 that allows extracting fuzzy rules from SVMs.

Parameters of membership functions can be optimized utilizing gradient decent algorithms of adaptive networks. In general, optimal fuzzy sets have different variances, while the kernel functions have the same ones. In order to obtain a set of optimal fuzzy with different variances, we can adopt methods such as gradient decent algorithms or GAs. We derive the following adaptive algorithm to update the parameters in the fuzzy membership functions:

$$\sigma_i(t+1) = \sigma_i(t) \delta \varepsilon_{1,i} \left[\frac{(x-c)^2}{\sigma^3} \exp\left(-\frac{(x-c)^2}{2\sigma^2} \right) \right] \quad (13)$$

$$c_i(t+1) = c_i(t) \delta \varepsilon_{1,i} \left[\frac{-(x-c)}{\sigma^2} \exp\left(-\frac{(x-c)^2}{2\sigma^2} \right) \right] \quad (14)$$

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

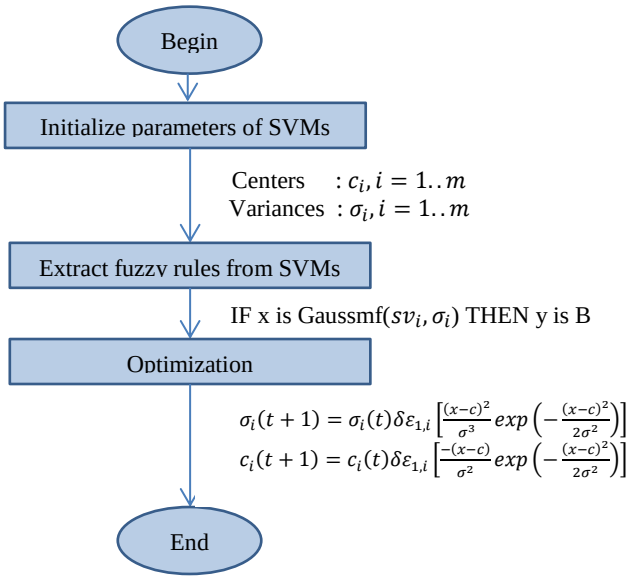


Figure 1. Block diagram of f-SVM algorithm.

III. Data Clustering using Self-Organizing Maps

SOM (Self-Organizing Map) is a type of artificial neural network that is trained by using unsupervised learning that was introduced by Kohonen [12], [18]. This model was proposed as an effective solution toward the recognition and control of robots. In SOM, the output neurons are usually organized into D-dimensional map in which each output neuron is connected to each input neuron. The arrangement of neuron is a regular spacing in a hexagonal or rectangular grid. The structure of a Kohonen SOM is shown in Figure 2.

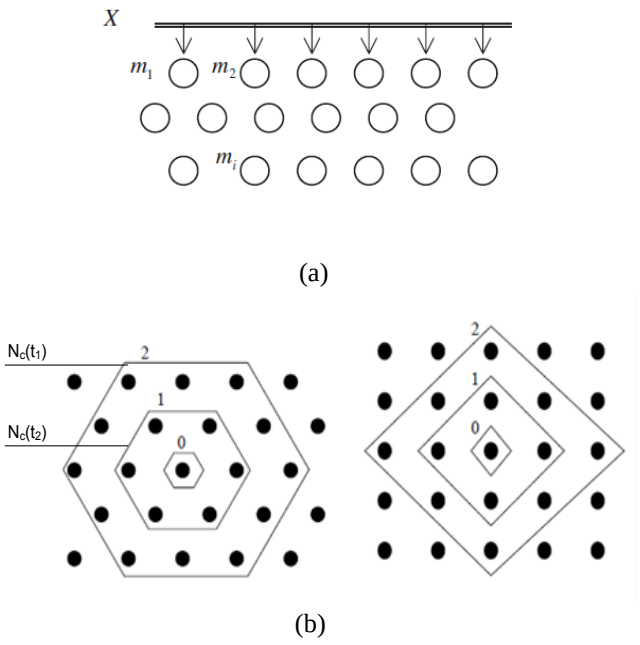


Figure 2. (a) An SOM example. (b) The distribution of rectangular and hexagonal SOM

As Figure 2, in SOM, each neuron is associated with a reference vector m_i and neighborhood range N_c. The reference vector has to be the same size as the size of the input vector and is used as the measure of closeness between input vectors. The neighborhood range is a symmetric function and also monotonically decreases with the distance between neurons in the map and centre neuron (wining neuron).

The SOM generalizes the wining neuron to its neighbors on the map by performing the training algorithm for the input vectors. The final result is that the neurons on the map ordered: neighboring neurons have similar weight vectors (Figure 2b). SOM is widely used for clustering because after training, the output neurons of SOM are automatically organized into a meaningful two-dimensional order in which similar neurons are closer to each other than the more dissimilar ones in terms of the reference vectors, thus keep close in the output space for the data points which are similar in the input space. Recent studies have suggested using SOM as quite an effective solution for stock market data [4], [11]. In this research, we do not have desire for in-depth analyses of machine learning SOM, the details of SOM have been presented in [12], [18]. Many researches in [4], [11] have also improved the effectiveness of combining SOM and SVMs model for data clustering both from theoretical and empirical analysis.

IV. The Stock Price Forecasting Model Based on Combination of SOM and f-SVM

In this research, the purpose to predict stock market and we propose a fuzzy inference model based on fuzzy rules extraction method from transaction history data. The model, which extracts fuzzy rules from data, is constructed by combining the cluster technique using SOM and f-SVM algorithm. A diagram of stock market prediction model is presented in Figure 3.

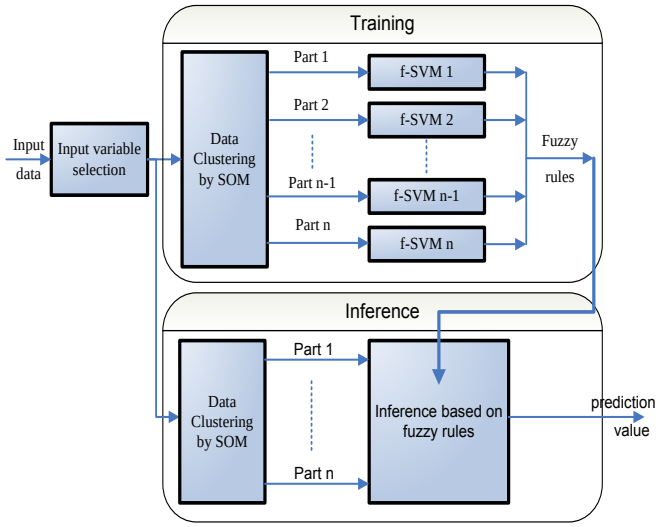


Figure 3. Block diagram of forecasting model.

A. Input variable selection

The results of other authors on stock market predictability showed that there are many ways to select input variables, such

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

as using daily stock market index <opening, high, low, closing price> [8], [17], macroeconomic indicators[1], ,… In this model, we have chosen stock market index as input variable.

According to the analysis and evaluation of L.J. Cao and Francis E.H. Tay in [8], 5-day relative difference in percentage of price - RDP is more effective, especially in the stock market prediction performance. In this model, we select the input variables based on the proposal and calculation of L.J. Cao and Francis E.H. Table 1 presents selected variables and their calculations.

TABLE I. TABLE TYPE STYLES

Symbol	Variable	Calculation
x_1	EMA100	$P_i - \overline{EMA_{100}(i)}$
x_2	RDP-5	$\big(P(i) - P(i-5)\big)/P(i-5) * 100$
x_3	RDP-10	$\big(P(i) - P(i-10)\big)/P(i-10) * 100$
x_4	RDP-15	$\big(P(i) - P(i-15)\big)/P(i-15) * 100$
x_5	RDP-20	$\big(P(i) - P(i-20)\big)/P(i-20) * 100$
y	RDP+5	$(\overline{P(i+5)} - \overline{P(i)})/\overline{P(i)} * 100$ $\overline{P(i)} = \overline{EMA_3(i)}$

where P(i) is closing price of the *i-th* day, and $EMA_m(i)$ is *m-day* exponential moving average closing price of the *i-th* day.

B. Clustering data by SOM

For data mining application, typically we work with a large volume data while many algorithms are ineffective for large data set. A common approach to solve this problem is split input data into smaller clusters, then apply the learning algorithm to each cluster and synthesize the results of simulation studies [13]. Moreover, one of the problems in financial time series forecasting is that time series are non-stationary. Statistics of stock prices depend on different factors such as economic growth and recession, political situation, environment, calamity… There is a limitation to find out stock price prediction rules based on historical market data. In the proposed model, the SOM is used to decompose the whole input space into regions where data points with similar statistical distributions are grouped together, so as to capture the non-stationary property of financial series.

The results of clustering of data by SOM provide an effective way to solve the two problems [4]: 1) Reducing data to a small number of dimensions is useful for increasing the speed of the model. 2) The data clusters are equivalent in statistical distributions to avoid interference.

C. Fuzzy rules extraction by f-SVM

Each cluster which was clustered by SOM will be trained for respective f-SVM machine to extract fuzzy rules. As shown in detail in the section 2, f-SVM machine extracts the fuzzy rules from each cluster of input data based on support vectors obtained from the SVM module which is integrated inside. By extracting fuzzy rules from SVM we will obtain rule sets in form:

$$IF\ x_1\ is\ Gaussmf\left(sv_1^i, \sigma_1^i\right) and\ x_2\ is\ Gaussmf\left(sv_2^i, \sigma_2^i\right)$$

$$and\ \dots\ x_j\ is\ Gaussmf\left(sv_j^i, \sigma_j^i\right) \dots\ THEN\ y\ is\ B^i$$

where $\mathrm{Gaussmf}\left(\mathrm{sv}_\mathrm{j}^\mathrm{i}, \sigma_\mathrm{j}^\mathrm{i}\right)$ is Gauss membership function.

D. Stock market prediction based on fuzzy rules

Extraction of fuzzy rules from f-SVM machine is an effective method for predicting stock market movements. Clustering low size data will reduce the complexity of fuzzy inference algorithm.

The above fuzzy rules in data mining have a certain distance to the understanding of human experts; however, fuzzy clustering is a condition for human expert to understand and evaluate these rules.

V. EXPERIMENT AND RESULTS

In order to evaluate the performance of the proposed model, we build a test system based on Matlab Toolkit. In this system, the SOM tool for Matlab is used to partition input data into several buckets, that toolbox developed by Esa Alhoniemi, Johan Himberg, Juha Parhankangas and Juha Vesanto [20]. The SOM Toolbox can be downloaded from http://www.cis.hut.fi/projects/somtoolbox/. To produce support vectors from training data we used LIBSVM – a library for Support vector Machines developed by Chih-Chung Chang and Chih-Jen Lin [19], which can be downloaded from http://www.csie.ntu.edu.tw/~cjlin/libsvm/. Finally, we use AVALFIS function in Matlab Fuzzy Logic Toolkit to infer stock market prediction based on producing fuzzy rules.

A. Data sets

The experimental data source was chosen from famous individual companies and composite indexes in America includes IBM Corporation stock (IBM), the Apple inc. stock (APPL), the Standard & Poor's stock index (S&P500), and the Down Jones Industrial Average index (DJI). All data used in this work are downloaded from Yahoo Finance http://finance.yahoo.com/

The daily data including Close-Price of four stocks are used as data sets for experimental. List of data sources are presented in Table 2. For each data set, the data is divided into two subsets according to the time sequence - training and testing subsets. With the objective of maximizing the size of training data to increase the coverage capability of training data samples, there are only 200 data samples used for the testing subset and all of the rest data are used for the training subset.

TABLE II. THE DATA SOURCE INFORMATION

Stock name	Time period	Training data	Testing data
IBM Corporation stock (IBM)	03/01/2000 - 30/06/2010	2409	200
Apple inc. stock (APPL)	03/01/2000 - 30/06/2010	2409	200
Standard & Poor's stock index (S&P500)	03/01/2000 - 23/12/2008	2028	200
Down Jones Industrial Average index (DJI)	02/01/1991 - 28/03/2002	2352	200

B. Performance metrics

The performance metrics used to evaluate in this study are the normalized mean squared error (NMSE), mean absolute error (MAE), and directional symmetry (DS) [4][8][11]. Among them, NMSE and MAE are measures of deviation between the actual value and the forecast value. DS provides an

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

indication of the predicted direction of RDP+5 given in the form of percentages. The predicted results are better if the values of NMSE and MAE are small, while large value of DS is better. The definitions of these metrics can be found in Table 3.

TABLE III. METRICS

Metrics	Calculation
NMSE	$\frac{1}{\sigma^2 n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$ $\sigma^2=\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2$ $\bar{y}=\sum_{i=1}^{n}y_i$
MAE	$\frac{1}{n}\sum_{i=1}^{n}\lvert y_i-\hat{y}_i\rvert$
DS	$\frac{100}{n}\sum_{i=1}^{n}d_i$ $d_i=\begin{cases}1 & (y_i-y_{i-1})(\hat{y}_i-\hat{y}_{i-1})\geq 0\\ 0 & otherwise\end{cases}$

n is the total number of data patterns

y and ŷ represent the actual and predicted output value

C. Experimental Results

Table 4 presents a group of fuzzy rules are produced from S&P500 stock data.

TABLE IV. A GROUP OF FUZZY RULES ARE PRODUCED FROM S&P500 STOCK DATA

Rule	Detail
R_1	IF x_1=Gaussmf(0.09,-0.11) and x_2=Gaussmf(0.09,-0.12) and x_3=Gaussmf(0.09,-0.04) and x_4=Gaussmf(0.09,-0.10) and x_5=Gaussmf(0.09,-0.09) THEN y=0.10
R_2	IF x_1=Gaussmf(0.10,-0.01) and x_2=Gaussmf(0.09,-0.06) and x_3=Gaussmf(0.10,0.04) and x_4=Gaussmf(0.10,-0.10) and x_5=Gaussmf(0.10,-0.12) THEN y=0.57
R_3	IF x_1=Gaussmf(0.09,0.02) and x_2=Gaussmf(0.10,0.02) and x_3=Gaussmf(0.09,0.08) and x_4=Gaussmf(0.10,-0.08) and x_5=Gaussmf(0.10,-0.13) THEN y=-0.02
R_4	IF x_1=Gaussmf(0.10,-0.04) and x_2=Gaussmf(0.10,-0.08) and x_3=Gaussmf(0.10,0.02) and x_4=Gaussmf(0.09,-0.08) and x_5=Gaussmf(0.09,-0.11) THEN y=-0.29
R5	IF x1=Gaussmf(0.10,-0.03) and x2=Gaussmf(0.09,-0.06) and x3=Gaussmf(0.10,0.03) and x4=Gaussmf(0.09,-0.10) and x5=Gaussmf(0.09,-0.13) THEN y=-0.38

We conduct an experiment to compare the results between the proposed model which predicts stock market based on fuzzy rules extraction and other models such as the RBN model and the hybrid model of SOM and SVM with the same testing data (200 samples). RBN model was built on a generalized regression neural network which is a type of Radial Basis Network (RBN). The generalized regression neural network is often used for prediction problems in [7], [14], [16]. The hybrid model of SOM and SVM was proposed to improving the effectiveness of time-series forecasting, especially stock market prediction [4], [11]. Moreover, we have compared with the experiment results of ANFIS model (Adaptive Neural Fuzzy Inference System). ANFIS model is a fuzzy neural network model which was proposed and standardized in Matlab. ANFIS has been applied in several studies in prediction problems. The prediction performance is evaluated using the following statistical metrics: NMSE, MAE, and DS. The results of the proposed model and other models are shown inTable 5&6.

TABLE V. RESULTS OF RBN AND SOM+ANFIS

Stock code	RBN			SOM+ANFIS		
	NMSE	*MAE*	*DS*	*NMSE*	*MAE*	*DS*
IBM	1.1510	0.0577	43.72	1.2203	0.0617	47.74
APPL	1.3180	0.0475	45.73	2.8274	0.0650	49.75
SP500	1.2578	0.1322	51.76	1.7836	0.1421	48.24
DJI	1.0725	0.1191	50.75	1.7602	0.1614	49.75

TABLE VI. RESULTS OF SOM+SVM AND SOM+F-SVM

Stock code	SOM+SVM			SOM+f-SVM		
	NMSE	*MAE*	*DS*	*NMSE*	*MAE*	*DS*
IBM	1.1028	0.0577	44.22	1.0324	0.0554	50.75
APPL	1.1100	0.0445	52.76	1.0467	0.0435	53.27
SP500	1.1081	0.1217	52.76	1.0836	0.1207	53.27
DJI	1.0676	0.1186	50.25	1.0459	0.1181	51.76

The experiment results in Table 5&6 demonstrates that the MNSE and MAE of SOM+f-SVM model are smaller than RBN and SOM+ANFIS, indicating that there is a smaller deviation between the actual and predict values in SOM+f-SVM. Moreover, the DS (Directional Symmetry) of the proposed model is higher than RBN and SOM+ANFIS. This shows that the predictions of SOM+f-SVM are more accurate than those of two other models. The comparison between the SOM+f-SVM model and SOM+SVM model (L.J. Cao and Francis E.H in [4]) is shown in Table 5, which shows that the values of MNSE, MAE and DS of the proposed model have not improvement significantly. This is obvious, because f-SVM algorithm used in proposed model perform extracts fuzzy rules from SVMs. The SOM+SVM model is used as "black-box" learning and inference processes. Otherwise, the proposed model allows producing a set of fuzzy rules and the inference processes will be performed using these rules. Results of learning process which is fuzzy rules extraction from SVMs have gradually clarified "black-box" model of SVMs. Based on the set of extracted rules, the human experts can understand and interact to improve the efficiency of using set of rules for inference process. In addition, using SOM for data clustering to split the input data into several smaller datasets brings the following effects: reducing the size of input data and thereby reducing the complexity of the algorithm, the generated rules will be split into several clusters, respectively. It helps human experts to understand and analysis fuzzy rules easily.

VI. CONCLUSIONS

In this study, we proposed an F-SVM algorithm to extract fuzzy rules from SVM; then we developed a stock market prediction model based on combination SOM and F-SVM. The experiment results showed that the proposed model has been used to predict stock market more effectively than the previous models, reflected in better values of three parameters: NMSE, MAE and DS. In addition, data cluster with SOM has been used to improve execution time of algorithms significantly in

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

this model. Otherwise, as shown in section 5.2 of this paper, the efficacy of the proposed model is the use of extraction of fuzzy rules which is a form of splitting set of rules; it helped in analyzing fuzzy rules easily. However, there are some drawbacks in SVM model: if it improves the accuracy of the model, the number of SVs will be increased; which causes an increase of the number of fuzzy rules. Thus, the system is more complex, especially the interpretability of the set of rules will decrease and then, human experts have difficulties in understanding and analyzing those rule sets.

In future work, we will concentrate on finding solutions to improve the interpretability of the sets of rules which are extracted from SVMs. After solving this problem, we gain the basis for analyzing the sets of rules and then optimize them in order to improve the effect of prediction.

REFERENCES

[1] Christan Pierdzioch, Jorg Dopke, Daniel Hartmann, "Forecasting stock market volatility with macroeconomic variables in real time," Journal of Economics and Business 60, 2008, 256-276.

[2] Corinna Cortes and Vladimir Vapnik, "Support-Vector Networks," Machine Learning, 20, 1995, 273-297.

[3] Hajizadeh E., Ardakani H. D., Shahrabi J. "Application Of Data Mining Techniques In Stock Markets: A Survey," Journal of Economics and International Finance Vol. 2(7), 2010, 109-118.

[4] Francis Eng Hock Tay and Li Yuan Cao, "Improved financial time series forecasting by combining Support Vector Machines with self-organizing feature map," Intelligent Data Analysis 5, IOS press 2001, 339-354.

[5] J.-H Chiang and P.-Y Hao, "Support vector learning mechanism for fuzzy rule-based modeling: a new approach," IEEE Trans. On Fuzzy Systems, vol. 12, 2004, pp. 1-12.

[6] J.L. Castro, L.D. Flores-Hidalgo, C.J. Mantas and J.M. Puche, "Extraction of fuzzy rules from support vector machines," Elsevier. Fuzzy Sets and Systems, 158, 2007, 2057 – 2077.

[7] Kreesuradej W., Wunsch D., Lane M. "Time-delay Neural Network For Small Time Series Data Sets," in World Cong. Neural Networks, San Diego, CA 1994.

[8] L.J.Cao and Francis E.H.Tay, "Support vector machine with adaptive parameters in Financial time series forecasting," IEEE trans. on neural network,vol. 14, no. 6, 2003.

[9] Md. Rafiul Hassan, Baikunth Nath, Michael Kirley, "A fusion model of HMM, ANN and GA for stock market forecasting," Expert Systems with Applications 33, 2007, 171–18

[10] S. Chen, J. Wang and D. Wang, "Extraction of fuzzy rules by using support vector machines," IEEE, Computer society, 2008, pp. 438-441

[11] Sheng-Hsun Hsu, JJ Po-An Hsieh, Ting-Chih CHih, Kuei-Chu Hsu, "A two-stage architecture for stock price forecasting by integrating self-organizing map and support vector regression," Expert system with applications 36, 2009, 7947-7951

[12] Teuvo Kohonen, "The self-organizing map," Elsevier, Neurocomputing 21, 1998, 1-6

[13] T.G. Dietterich, "Machine learning research: Four current directions,' AI Magazine, 18(4), 1997, 97-136

[14] Younes Chtioui, Suranjan Panigrahi, Leonard Francl, "A generalized regression neural network and its application for leaf wetness prediction to forecast plant disease," Chemometrics and Intelligent Laboratory System 48, 1999, 47-58

[15] R. Courant, D. Hilbert, Methods of Mathematical Physics. Wiley, New York (1953)

[16] Iffat A. Gheyas, Leslies S. Smith. "A Neural network approach to time series forecasting," Proceeding of the World congress on Engineering 2009, Vol II

[17] Md. Rafiul Hassan and Baikunth Nath."Stock market forecasting using Hidden markov model: A new approach," 5th International conference on intelligent system design and applications (ISDA'05), 2005

[18] Teuvo Kohonen, "The self-organizing map," Proceeding of The IEEE, Vol. 78, No. 9, 1990

[19] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen lin, "A practical Guide to Support Vector Classification," http://www.csie.ntu.edu.tw/~cjlin/libsvm/, 2010

[20] Juha Vesanto, Johan Himberg, Esa Alhoniemi, and Jaha Parhankangas, SOM Toolkox for Matlab 5, http://www.cis.hut.fi/projects/somtoolbox/, 2000

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Limitations of Current Security Measures to Address Information Leakage Attacks

Omar Hussein [1], Nermin Hamza [2], Hesham Hefny [3]

Computer and Information Sciences Department
Institute of Statistical Studies and Research, Cairo University, Egypt
[1] ohusseins@gmail.com
[2] nermin.hamza@cu.edu.eg
[3] hehefny@ieee.org

***Abstract*—Information leakage attacks represent a serious threat for their widespread and devastating effects. Their significance stems from the fact that they are committed by an organization's authorized computer users, and/or processes executing on their behalf. The diverse avenues that could be exploited to carry out such attacks add another barrier towards addressing them. Based on literature review, this paper explores strengths of security measures intended to confront information leakage attacks, and focuses on pinpointing their respective limitations. It demonstrates that only few of them are capable of mitigating such attacks, whereas the rest suffer from conceptual and/or implementation-related limitations that render them vulnerable to circumvention. They are basically prone to high false positive and/or false negative rates, complex to apply, inflexible during execution, suffer from degraded performance, or require hardware modification. Most importantly, neither of them provides a remedy for new undetected malicious software, nor the ever increasing insider threat.**

***Index Terms*—Information Security, Information Leakage, Security Measures, Security Limitations**

I. Introduction

Confidential information such as trade secrets, design documents, business plans, and customer information constitute the main assets of an organization. Ensuring confidentiality preservation of such critical assets whilst controlling their propagation is of paramount importance for an organizations stability and growth. Conversely, information leakage means unauthorized information propagation; making private information public. It refers to incidents where restrictions on information propagation have been breached, and confidential information has been revealed to unauthorized parties.

In recent years information leakage attacks have gained extensive concern from the information security community. Despite these valuable efforts that have been made so far, information leakage remains an unsolved problem, and is still prevailing. Damaged business reputation that could be measured by sharp sales declines, and negative reactions to an organizations image are examples of probable consequences to information leakage attacks. Prominent examples include [43] and [33].

Significance of information leakage stems from the fact that at time of attack, adversaries had authorized access to confidential data, and the adopted security measures were considered sufficient. Furthermore, the scope of information leakage is very broad, where multiple avenues to leak out information could be exploited. It could occur via e-mails from work, removable storage media, instant messaging, cloud computing, remote network access, hacker break-ins, etc. This paper investigates security measures that are currently adopted to defend against information leakage attacks. Focus is mainly driven towards malicious software (malware) and the insider threat for being the most prominent perpetrators of such attacks.

Malware is responsible for 80% of data loss incidents, as reported in the 2011 Data Breach Investigations Report [40]. For instance, spyware programs record keystrokes including usernames and passwords in order to open backdoors for attacks to be mounted. They could also send accessible data over the network. New malware poses a major threat and is of particular interest for its ability to evade detection by current malware detection methods, and for being released at a rate of tens of thousands a day. In the first half of 2010, Sophos [34] received around 60,000 new malware samples every day, and in the first quarter of 2011, 73,000 new malware were identified on average every day [24].

Insiders can leak out confidential information for financial gain or to future employers. Their threat is prevailing; according to the 2010 Data Breach Investigations Report [39], 48% of data breaches were caused by insiders. A joint survey conducted by Symantec corporation and Ponemon Institute revealed that around 60% of employees who were fired, or resigned from their jobs have been accused of leaking out their companies' confidential information [36]. The insider threat represents a challenging unsolved problem for three main reasons: (1) at time of attack, insiders had authorized access to victim systems; (2) they possess deep understanding of the targeted vulnerable processes; and (3) they are aware of systems' unpatched security vulnerabilities. The remainder of this paper is organized as follows: Section II investigates security measures that are currently used to defend against information leakage attacks. It illustrates their strengths and focuses on pinpointing their respective limitations. Finally, Section III concludes the paper.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

II. RELATED WORK

Several security measures are currently used to confront information leakage attacks: (1) malware detection; (2) access control; (3) encryption; (4) digital rights management; (5) decentralized information flow control; (6) information flow tracking; (7) system call monitoring; (8) data loss prevention systems; and (9) hardware-based approach. The following subsections discuss each of these security measures to illustrate their respective strengths and weaknesses.

A. Malware Deteciton

Malware detection methods can be generally categorized into two classes: signature-based and heuristics-based methods. Signature-based is the most commonly used method to detect malware. It is reactive in the sense that malware needs to be analyzed first prior to executing the detection and disinfection procedures. Consequently, the method's effectiveness totally depends on how rapid new malware is analyzed, and its signature identified and incorporated into the anti-virus software database.

However, signature-based malware detection method is prone to false negatives (i.e., missed detections) for two main reasons. Firstly, by packing and using code obfuscation techniques, attackers could transform previously known malware to new unknown ones [13]. Given the fact that the majority of new malware that appear nowadays are packed versions of known ones, therefore, they can easily evade detection. They were not analyzed, and their signatures were not included in the anti-virus database [10].Secondly, with the wide spread of crimeware toolkits and being increasingly profitable, creation and release of new sophisticated malware became easier even for casual attackers [7]. According to the 2011 Data Breach Investigations Report [40], 63% of malware security breaches were committed by highly customized malware. Consequently, the method's ad-hoc nature makes it vulnerable to circumvention; it can neither detect new malware nor packed versions of known ones.

Heuristics-based malware detection method is often based on comparing malware against a set of rules to determine whether a file should be deemed suspicious [41]; for instance, by monitoring modifications to the system registry and/or certain libraries. However, this method suffers from two limitations: (1) it yields high false positive rates. A benign application is considered compromised just because it modified the system registry during its installation procedure; and (2) it yields high false negative rates. Malware may either attempt to modify libraries that detectors do not monitor, or could hide its entries in the system registry, as in case of kernel-mode rootkits [9]. As a reflection of the aforementioned shortcomings, current malware detection rates are shocking. Popular anti-virus tools can detect on average less than 19% of newly released malware [12].

B. Access Control

Access control aims at protecting systems resources from unauthorized access. It specifies which resources could be accessed in which access mode by whom of the identified principals, and how they share data with others [4]. Access control security policies (ACSPs) are classified into discretionary access control (DAC), MAC [32], and role-based access control (RBAC) [14].

In DAC objects' owners specify upon their own individual discretion allowed access permissions to subjects. However, DAC is criticized for being defenseless against malware attacks and the insider threat for three main reasons [19]. Firstly, it is transitive; an insider who was granted read access permission on some file can illegally pass his/her privilege along to other unauthorized user(s) by copying the file's contents to an object under his/her control. Secondly, programs inherit the identity and access permissions of the executing user. This enables malware to circumvent DAC access restrictions. For example, an insider can bypass DAC restrictions by giving an accountant a Trojan-horsed program that apparently performs a desired function. However, the program, during its course of execution, covertly copies the accountant's confidential files to the insider's desired location. Investigating audit trails would indicate that the accountant leaked out his/her confidential files. Thirdly, objects' owners maintain their own individual security policies. Consequently, a centralized uniform organization-wide ACSP is unenforceable.

MAC aims at resolving DAC limitations. Its implementation is called multi-level security. It enforces a centralized uniform system-wide ACSP in which: (1) objects and subjects in the system are classified into security levels (SLs); (2) access is allowed or denied according to a set of fixed rules, where SLs associated with objects are compared with those associated with subjects; and (3) objects' owners cannot alter access permissions except through a central authority (e.g., the security administrator). Such controlled information flow enables containing and mitigating malware and insiders ability to leak out confidential information. However, multi-level security suffers from a fundamental limitation; it is too strict and inflexible during execution. In addition, correct policy configuration is complicated, which hinders its applicability [16].

RBAC enables specification, enforcement, and management of a central organization-wide ACSP. In RBAC objects' owners do not specify access permissions upon their own discretion. Instead, the security administrator creates roles; each representing a job, a function, or a position in the organization. The security administrator associates access permissions and users with roles. Consequently, users acquire access permissions associated with their assigned roles. RBAC simplifies authorizations management, and provides flexibility in specifying and enforcing a central ACSP. However, it suffers from two main limitations [19]: (1) it assumes that all access

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

permissions needed to accomplish a certain job in an organization can be precisely identified and included in a single role. Unfortunately, this is not always practically feasible; and (2) to fulfil the *separation of duties* security principle, mutually exclusive roles needed to complete every critical task should be created, and each role should be cautiously associated only with the relevant participating user. However, this can never be guaranteed; the security administrator can intentionally or inadvertently breach this principle by associating all mutually exclusive roles needed to perform a critical task with a single user.

C. Encryption

Encryption is the process of converting the original message (called plaintext) into a coded message (called ciphertext). The goal is to ensure that confidential data is disclosed only to the intended recipients. Decryption is the reversible process in which plaintext is restored from the ciphertext [35]. Relying solely on encryption is insufficient to address information leakage attacks for two main conceptual reasons: (1) its ultimate objective is to forbid disclosure of confidential information except for secret key holders. However, after decryption, no defenses are available to defend against information leakage attacks; and (2) leaking out the secret key to unauthorized users disables the whole desired protection.

D. Digital Rights Management

Digital rights management (DRM) systems [47] combine access control with encryption to extend protection to digital contents beyond restricting their accessibility, to controlling their usage post-distribution outside of the organization's network perimeter. In a DRM system a digital content is encrypted and packaged into a content object for distribution. To decrypt and access the content object, the user needs a license (also called the rights object). It contains the decryption key and the access rights the user has acquired for that content object. Licenses are usually stored in, and distributed from, a separate server apart from that of the corresponding content objects. Users' acquired access rights are interpreted and enforced by the DRM software running on the client side.

To protect licenses from illegal sharing, they are usually locked to users' computers. Consequently, unlicensed users are prohibited from gaining access to content objects even if they have both the objects and the licenses. In addition, licensing eliminates the need to directly contact objects' owners; thereby, reducing security administration overhead. Furthermore, DRM provides additional and finer-grained usage restrictions imposed on content objects, as compared to those traditionally specified in ACSPs. For example, it allows controlling number of allowed printouts, and providing dynamic watermarking to determine the source of printed copies. The latter restriction is extremely important to enable identification and prosecution of pirates [25].

However, DRM systems suffer from three main limitations: (1) poor portability of content objects' across multiple devices due to DRM's reliance on device-based rather than user-based authentication, (2) inflexibility in modifying access rights set for content objects, where they merely revoke entire devices instead of managing access rights set for individual objects [5]; and (3) most document security DRM vendors, such as [2], implement their controls in the form of plug-ins in specific applications (i.e., Microsoft Office) running above a specific operating system (OS) (i.e., Microsoft Windows). However, it cannot be always assumed that users are limited to these specific software packages.

E. Decentralized Information Flow Control

As explained earlier, DAC and RBAC suffer from lack of control over information flow, whereas in MAC information flow directions are predetermined and static. In this context, [28] present their decentralized information flow control (DIFC) model, in which MAC rigid restrictions are extended. Users are allowed in a decentralized way to restrict information flow by associating security labels with data that they own. The model defines a set of rules to share information with untrusted code (e.g., applets) to control how this code disseminates information. DIFC can be supported at either of the programming, OS, or architecture level as follows:

- *Programming Level*

Java information flow (JIF) [29] is a language-level implementation of DIFC. It applies the DIFC model as an extension to Java programming language. By associating security labels with language-level variables, JIF enables programmers to control data flows within a program, and between the program and the outside world. However, JIF suffers from three significant conceptual limitations: (1) applications are left wide open for insiders to carry out any malicious manipulation, where it is up to the programmers to specify information flow security policies (IFSPs). Users are forced to trust them, whereas their own preferences are ignored; (2) IFSPs are complex to apply, where programmers should associate a security label with every variable. This could introduce security-related errors, especially when developing distributed systems [1]; and (3) IFSPs are individually specified and are enforced on ad-hoc basis. This leads to inconsistency in application, and increased security administration overhead.

- *OS Level*

Flume [23] is an OS-level implementation of DIFC in the form of user-level reference monitor for Linux OS. It enforces DIFC policy at the granularity of processes. Its interface helps programmers secure existing applications and write new ones with current tools and libraries. However, Flume suffers from three main limitations: (1) as any software executing above an OS, it is prone to Linux's security-related vulnerabilities that could result in gaining *super-user* privileges; (2) it is of limited applicability for its complexity, where it requires developers to partition each application into a number of components, each with a different set of privileges; and (3) it results in a significant performance penalty whilst interposing system

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

calls.

Other researchers proposed completely new DIFC-based OSs, such as HiStar [45], which enforces DIFC policy at the granularity of files and processes. It aims at containing exploitable software vulnerabilities through isolating applications code, and controlling inter-process information flow. DStar [46] extends HiStar to work in a network to mitigate effects of untrustworthy distributed applications. For each machine, DStar adds an *exporter daemon*, which is responsible for restricting information flow between any two processes running on two different connected machines.

These OSs are advantageous in being transparent to developers. However, they are constrained by two main limitations that prevent their applicability: (1) they are experimental; legacy applications cannot run above them, and they need to be rewritten, which is unreasonable; and (2) they suffer from degraded performance, where every single inter-process communication is verified [1].

- *Architecture Level*

The RIFLE architecture [38] aims at enforcing run-time user-defined IFSPs. To control information flow, RIFLE proposes a combination of modified hardware architecture to enable assignment of security labels to memory words, along with program *binary translation*. However, RIFLE is constrained by a significant conceptual limitation; it requires hardware modification, which makes it inapplicable to existing machines.

F. Information Flow Tracking

Processing of confidential data may propagate secrets to unauthorized parties. In this context, [30] proposed their information flow tracking approach called dynamic taint analysis (DTA). It is based on the claim that security exploits occur because applications accept input from untrusted sources. It attempts to track and restrict information flow by tagging (tainting) data from potentially untrusted input channels (e.g., networks) during program execution. To ensure that untrusted input will not violate the intended IFSP, it is validated. Events such as, incorrect validation, or bypassing a validation check may indicate an attack. After detection, DTA raises a flag, and stops the program execution. Examples include: TaintTracker [30], and TaintEraser [48], which allows users to track sensitive input (e.g., entered passwords). However, it enforces an IFSP only on a single machine, not over a network of machines.

Unfortunately, DTA suffers from a significant limitation; it is prone to false negatives caused by the *under-tainting problem* [20]. DTA tracks only explicit data flows; it ignores tracking control flows that leads to implicit data flows. That is, DTA overlooks the fact that tagged data values may affect programs control flow that in turn may affect other data values. As a result, all values that should be tagged are not; thereby, causing information leakage detection failure. To reveal conditions that cause this problem, DTA needs to explore all possible execution paths in a program by analyzing its source code.

G. System Call Monitoring

An application can be characterized by its application programming interface (API), and sequences of invoked system calls. The API represents the interface between applications processes and the OS. API system calls allow interacting with the OS, and accessing and modifying other objects in the system. Monitoring applications' sequences of invoked system calls enables observing their behavior. Christodorescu et al. [11] explain a system call tracing technique to extract sequences of system calls present in a set of benign applications, and those found in malware samples. The extracted system call sequences represent the execution behavior of these benign applications, and the semantic signature of those malware samples. This technique then compares both execution behaviours for malware detection purposes.

Malware and vulnerability analysts observed a common property in majority of application-level attacks; they do not conform to applications' expected normal behavior (i.e., they are anomalous / malicious). Host-based anomaly detection systems (HBADSs) rely on system call monitoring to determine whether applications have been compromised [15]. A HBADS constructs a unique normal behavior model for each application chosen to be monitored. It interposes between the monitored application and the underlying OS. It compares between the application's invoked system calls and its normal behavior profile. In case it detects a non-conforming system call invocation, the call is flagged as anomalous; the system raises an alarm, and stops the application execution to prevent adversaries from delivering their malicious payload.

Several approaches have been proposed to derive and construct models of applications' normal behavior. These approaches can be grouped into three categories: (1) monitor audit trails [22]; (2) static analysis of applications' source code [17] or of their binaries [21]; or (3) analyze applications' runtime behavior [27]. As shown in Figure 1, the solid line from the system call monitor through the system call interface to the OS kernel illustrates allowed system calls that have been checked by the system call monitor. The dashed line from the system call monitor indicates a call that has been flagged as anomalous because it did not conform to the application's normal behavior profile.

The main advantage of HBADSs is their ability to identify previously unknown attacks, even if adversaries gained control over the monitored applications. This is due to the attackers' ignorance of the differences between sequences of invoked system calls that conform to applications' normal behavior profiles, and those that do not. Nevertheless, HBADSs suffer from six limitations. Firstly, they are prone to a high rate of false positives caused by incomplete and/or incorrect modelling of applications' normal behavior [31].

Secondly, they may suffer from a high rate of false negatives caused by *mimicry attacks* that imitate legitimate system call sequences, whilst actually performing malicious actions [8]. For example, web server software attempts to open a

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

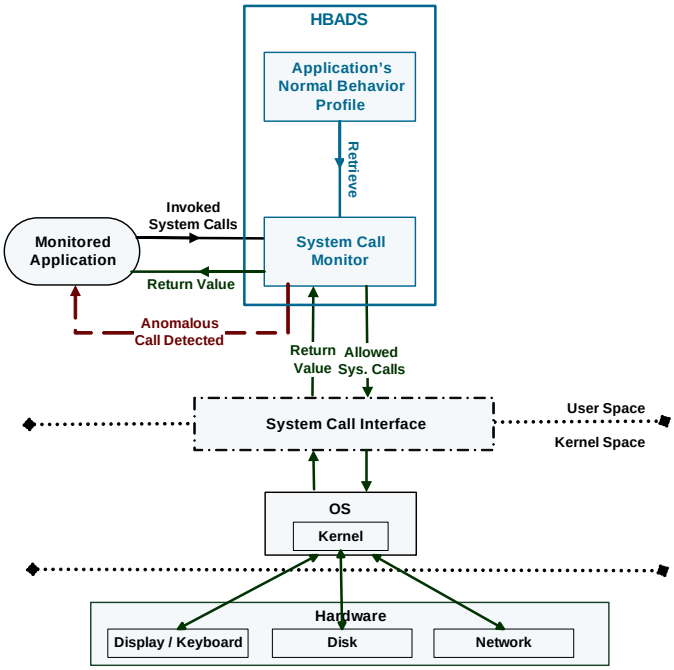


Fig. 1. A high-level view of a typical HBADS

file to send its contents over the network. Being undetected by the HBADS, a mimicry attack keeps the *open* system call unchanged. However, instead of passing the path to the intended file, it passes the path to a sensitive file instead; thereby, leaking confidential information over the network.

Thirdly, effectiveness of HBADSs could be limited in case static analysis of applications' source code was adopted to derive and construct models of their normal behavior. Attackers may exploit code paths in applications' source code that are not, or are scarcely used during the course of their execution [15]. Fourthly, it is assumed that applications' system calls are fixed. Though, this assumption does not always hold, as in the case of *self-modifying code*, which is used as an additional layer of complexity to make attacks more difficult [26]. Fifthly, attacks could potentially be incorporated into applications' normal behavior profiles in case their run-time behaviors were analyzed after being compromised. Sixthly, if a monitored application was reconfigured, then its normal behavior profile should be updated.

H. Data Loss Prevention Systems

Data loss prevention (DLP) systems are designed to prevent information leakage. In this context, Yumerefendi et al. proposed TightLip [44]. Through this system users tag their confidential data, such that, when a process receives a confidential input, a replicated process is created and prohibited from accessing it. Outputs from the original and the replicated processes are compared. If they were not identical (e.g., a network communication was established), then this indicates a potential attempt to leak out confidential information. However, TightLip suffers from two limitations: (1) it is prone to a high rate of false positives since none identical inputs often affect applications control flow causing none identical outputs; and (2) it is prone to a high rate of false negatives, as it identifies sensitive files by an ad-hoc combination of their type, path, and performing pattern-matching with their contents.

Commercially, several DLP systems, such as [37] have been developed to enforce central data leakage prevention policies. These systems basically rely on three techniques: (1) keyword matching, which is used in cases where few keywords can identify confidential data (e.g., medical and financial data); (2) regular expressions, which are suitable for well-structured data that exist in defined formats (e.g., phone numbers and addresses); and (3) hash fingerprinting, which is used for unstructured data that does not conform to defined formats (e.g., source code and design documents). In the latter technique, the DLP system populates a database with hashes of substrings derived from a given set of sensitive files. These hashes can then be used to identify other sensitive files.

However, DLP systems are confined by two main limitations: (1) a keyword list cannot always be accurately determined. It is limited by the ability to wrap up contents of a sensitive file in few words or phrases; and (2) regular expressions and hash fingerprinting can offer only limited protection against information leakage attacks. They are prone to false negatives; they cannot block deliberate information leakage attempts. For example, an experienced insider could easily circumvent these systems by intentionally rephrasing, reformatting, or encrypting sensitive files. He/she could then leak them out without being detected and/or blocked [18].

Content filters, such as Websense [42] aim at preventing network-based information leaks. They limit where hosts can send data by blocking access to classified lists of websites. However, they are prone to false negatives; experienced insiders can post sensitive information on public websites that receive input (e.g., Wikipedia), and later display it from elsewhere.

I. Hardware-Based Approach

Alawneh and Abbadi [3] introduced a mechanism to protect shared information among collaborating organizations via trusted platform module (TPM). By creating domains for TPM equipped devices, sensitive files that belong to a source organization can be accessed exclusively by the allowed devices in the destination organization according to the source organization's policy. However, this mechanism suffers from a conceptual limitation; the TPM is effective only in providing security guarantees during software load-time. It is ineffective in defending against malware attempts to compromise applications after being loaded in memory and during execution [6].

III. CONCLUSION

This paper demonstrated that only few security measures that are intended to defend against information leakage attacks have been able to mitigate such attacks, whereas the rest suffer from conceptual and/or implementation-related limitations that render them ineffective and vulnerable to circumvention. As

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

explained, this was basically attributed to their ad-hoc nature in the sense that they do not provide in-depth protection under various attack scenarios. The most common limitation was being prone to high false positive and false negative rates. This was prominent in malware detection methods, system call monitoring, and DLP systems. Information flow tracking suffers fundamentally from false negatives caused by the *under-tainting problem*.

Encryption is essential to forbid disclosure of confidential information except for secret key holders. However, after decryption, plaintext is completely vulnerable to information leakage attacks. In addition, leaking out the secret key to unauthorized users disables the whole desired protection. ACSPs, excluding MAC, suffer from lack of control over information flow; therefore they are ineffective against information leakage attacks. Conversely, in MAC, information flow directions are predetermined and static, thereby allowing containment of such attacks. Unfortunately, MAC suffers from hindered applicability for its complicated policy configuration, and inflexibility during execution.

DRM extends protection to digital contents beyond restricting their accessibility, to controlling their usage post-distribution outside of the organization's network perimeter. However, it suffers mainly from poor portability across multiple devices, and inflexibility in modifying access rights set for content objects. DIFC at the programming level, besides being complex to apply, leaves applications wide open for insiders to carry out any malicious manipulation, where it is up to the programmers to specify IFSPs. DIFC at the OS level is of limited applicability for its complexity, besides resulting in a significant performance penalty. DIFC at the architecture level requires hardware modification, which makes it inapplicable to existing machines. Finally, the hardware-based approach is ineffective in defending against malware attempts to compromise applications after being loaded in memory and during execution. Most importantly, neither of these security measures provides a remedy for new undetected malware, nor is equipped to defend against the ever increasing insider threat.

REFERENCES

[1] T. Abdellatif, L. Sfaxi, R. Robbana, and Y. Lakhnech, "Automating Information Flow Control in Component-based Distributed Systems," *in Proc. of the 14th International ACM Special Interest Group on Software Engineering (SIGSOFT) Symposium on Component Based Software Engineering*, Jun. 2011, pp. 73-82

[2] Adobe (2013): Digital Rights Management: To Share and Protect. [Online]. Available: http://www.adobe.com/manufacturing/resources/drm/

[3] M. Alawneh, and I. M. Abbadi, "Preventing Information Leakage between Collaborating Organisations,"*in Proc. of the 10th International Conference on Electronic Commerce*, 2008, pp. 1-10

[4] R. Anderson, *Security Engineering: A Guide to Building Dependable Distributed Systems*, Indiana, Wiley Publishing, 2008

[5] A. Arnab, and A. Hutchison, "An Evaluation Framework for DRM,"*in Proc. of the 6th International Workshop for Technical, Economic and Legal Aspects of Business Models for Virtual Goods, incorporating the 4th International Open Digital Rights Language (ODRL) Workshop*, Oct. 2008, pp. 176-200

[6] S. Balfe, E. Gallery, C. J. Mitchell, and K. G. Paterson, "Challenges for Trusted Computing,"*IEEE Security & Privacy Magazine*, vol. 6, no. 6, pp. 60-66, 2008

[7] H. Binsalleeh, T. Ormerod, A. Boukhtouta, P. Sinha, A. Youssef, M. Dbbabi, and L. Wang, "On the Analysis of the Zeus Botnet Crimeware Toolkit," *in Proc. of the 8th International Conference on Privacy, Security and Trust*, Aug. 2010, pp. 31-38

[8] M. D. Bond, V. Srivastava, K. S. McKinley, and V. Shmatikov, "Efficient, Context-Sensitive Detection of Real-World Semantic Attacks,"*in Proc. of the 5th ACM Workshop on Programming Languages and Analysis for Security*, Jun. 2010, pp. 1-11

[9] P. Bravo, and D. F. Garcia, "Proactive Detection of Kernel-Mode Rootkits,"*in Proc. of the 6th International Conference on Availability, Reliability and Security*, Aug. 2011, pp. 515-520

[10] J. Caballero, C. Grier, C. Kreibich, and V. Paxson, "Measuring Pay-per-Install: The Commoditization of Malware Distribution,"*in Proc. of the 20th USENIX Security Symposium*, Aug. 2011, pp. 187-202

[11] M. Christodorescu, S. Jha, and C. Kruegel, "Mining Specifications of Malicious Behavior,"*in Proc. of the 6th joint meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on the Foundations of Software Engineering*, Sep. 2007, pp. 5-14

[12] Cyveillance. (2010) Cyveillance testing finds AV vendors detect on average less than 19% of malware attacks. [Online]. Available: http://www.cyveillance.com/web/blog/press-release/cyveillance-testing-finds-av-vendors-detect-on-average-less-than-19-of-malware-attacks

[13] A. P. Czarnowski, "Reversing Python Objects,"*Virus Bulletin*, pp. 13-17, Jul. 2011

[14] D. Ferraiolo, D. Kuhn, and R. Chandramouli, *Role-Based Access Control*, 2nd ed., Boston, Artech House, 2007

[15] S. Forrest, S. Hofmeyr, and A. Somayaji, "The Evolution of System-Call Monitoring,"*in Proc. of the 24th Annual Computer Security Applications Conference*, Dec. 2008, pp. 418-430

[16] L. Franco, T. Sahama, and P. Croll, "Security Enhanced Linux to Enforce Mandatory Access Control in Health Information Systems,"*in Proc. of the 2nd Australasian Workshop on Health data and Knowledge Management*, Jan. 2008, pp. 27-33

[17] R. Gopalakrishna, E. H. Spafford, and J. Vitek, "Efficient Intrusion Detection Using Automaton Inlining,"*in Proc. of IEEE Symposium on Security and Privacy*, May 2005, pp. 18-31

[18] M. Hart, P. K. Manadhata, and R. Johnson, "Text Classification for Data Loss Prevention,"*in Proc. of the 11th International Symposium on Privacy Enhancing Technologies*, May 2011, pp. 18-37

[19] V. C. Hu, D. F. Ferraiolo, and D. R. Kuhn, "Assessment of Access Control Systems". National Institute of Standards and Technology (NIST), Interagency Report 7316, 2006

[20] M. G. Kang, S. McCamant, P. Poosankam, and D. Song, "DTA++: Dynamic Analysis with Targeted Control-Flow Propagation,"*in Proc. of the 18th Annual Network and Distributed System Security Symposium*, Feb. 2011, pp. 269-282

[21] E. Kirda, C. Kruegel, G. Banks, G. Vigna, and R. A. Kemmerer, "Behavior-Based Spyware Detection,"*in Proc. of the 15th USENIX Security Symposium*, Jul. 2006, pp. 273-288

[22] C. Ko, G. Fink, and K. Levitt, "Automated Detection of Vulnerabilities in Privileged Programs By Execution Monitoring,"*in Proc. of the 10th Annual Computer Security Applications Conference*, Dec. 1994, pp. 134-144

[23] M. Krohn, A. Yip, M. Brodsky, N. Cliffer, M. F. Kaashoek, E. Kohler, R. Morris, and M. Csail, "Information Flow Control for Standard OS Abstractions,"*in Proc. of the 21st ACM Symposium on Operating Systems Principles*, Oct. 2007, pp. 321-334

[24] M. Lennon (2011) PandaLabs: Over 5 Million New Malware Samples in Q3. [Online]. Available: http://www.securityweek.com/pandalabs-over-5-million-new-malware-samples-q3

[25] Locklizard (2013) Information & Data Leakage: Why DRM is required for preventing information leakage. [Online]. Available: http://www.locklizard.com/information_leakage.htm

[26] N. Mavrogiannopoulos, N. Kisserli, and B. Preneel, "A Taxonomy of Self-Modifying Code for Obfuscation," *Elsevier Computers & Security*, vol. 30, no. 8, pp. 679-691, Nov. 2011

[27] D. Mutz, F. Valeur, C. Kruege, and G. Vigna, "Anomalous System Call Detection,"*ACM Transactions on Information and System Security*, vol. 9, no. 1, pp. 61-93, 2006

[28] A. C. Myers, and B. Liskov,"A Decentralized Model for Information Flow Control,"*in Proc. of the 16th ACM Symposium on Operating Systems Principles*, Oct. 1997, pp. 129-142

[29] A. C. Myers, and B. Liskov, "Protecting Privacy Using the Decentralized Label Model," *ACM Transactions on Software Engineering and Methodology*, vol. 9, no. 4, pp. 410-442, 2000
[30] J. Newsome, and D. Song, "Dynamic Taint Analysis for Automatic Detection, Analysis, and Signature Generation of Exploits on Commodity Software,"*in Proc. of the 12th Network and Distributed System Security Symposium*, Feb. 2005
[31] W. Robertson, F. Maggi, C. Kruegel, and G. Vigna, "Effective Anomaly Detection with Scarce Training Data,"*in Proc. of the 17th Annual Network and Distributed System Security Symposium*, Feb. 2010
[32] R. S. Sandhu, and P. Samarati, "Access Control: Principles and Practice,"*IEEE Communications Magazine*, pp. 40-48, Sep. 1994
[33] M. Smith. (2013) NSA leaker comes forward, warns of agency's 'existential threat'. [Online]. Available: http://edition.cnn.com/2013/06/09/politics/nsa-leak-identity/index.html
[34] Sophos Inc. (2010) Security Threat Report: Mid-year 2010. [Online]. Available: http://www.sophos.com/en-us/medialibrary/Gated%20Assets/white%20papers/sophossecuritythreatreportmidyear2010wpna.pdf
[35] W. Stallings, *Cryptography and Network Security: Principles and Practices*, 5th ed., Boston, Pearson Education, 2011
[36] Symantec Corp. (2009) Press Release: More Than Half of Ex-Employees Admit to Stealing Company Data According to New Study. [Online]. Available: http://www.symantec.com/about/news/release/article.jsp?prid=20090223_01
[37] Symantec Corp. (2013) Symantec Data Loss Prevention: Data Leak Prevention. [Online]. Available: http://www.symantec.com/theme.jsp?themeid=dlp-family
[38] N. Vachharajani, M. J. Bridges, J. Chang, R. Rangan, G. Ottoni, J. A. Blome, G. A. Reis, M. Vachharajani, and D. I. August, "RIFLE: An Architectural Framework for User-Centric Information-Flow Security,"*in Proc. of the 37th Annual IEEE/ACM International Symposium on Microarchitecture*, December 2004, pp. 243-254
[39] Verizon RISK Team and U.S. Secret Service. (2010) Data Breach Investigations Report. [Online]. Available: http://www.verizonbusiness.com/resources/reports/rp_2010-data-breach-report_en_xg.pdf
[40] Verizon RISK Team, U.S. Secret Service, and the Dutch High Tech Crime Unit. (2011) Data Breach Investigations Report. [Online]. Available: http://www.verizonenterprise.com/resources/reports/rp_data-breach-investigations-report-2011_en_xg.pdf
[41] Virus BULLETIN. (2014) Heuristics: Malware detection method using rules and pattern-matching. [Online]. Available: nolinkurl-http://www.virusbtn.com/resources/glossary/heuristics.xml
[42] Websense, Inc. (2012) Web Security Suite. [Online]. Available: http://www.websense.com/content/WebSecurity.aspx
[43] WikiLeaks. (2013) http://wikileaks.org/
[44] A. R. Yumerefendi, B. Mickle, and L. P. Cox, "TightLip: Keeping applications from spilling the beans,"*in Proc. of the 4th USENIX Symposium on Networked Systems Design and Implementation*, April 2007, pp. 159-172
[45] N. Zeldovich, S. B. Wickizer, E. Kohler, and D. Mazieres, "Making Information Flow Explicit in HiStar,"*in Proc. of the 7th USENIX Symposium on Operating Systems Design and Implementation*, November 2006, pp. 263-278
[46] N. Zeldovich, S. B. Wickizer, and D. Mazieres, "Securing Distributed Systems with Information Flow Control,"*in Proc. of the 5th USENIX Symposium on Networked Systems Design & Implementation*, Apr. 2008, pp. 293-308
[47] W. Zeng, H. Yu, and C. Lin,*Multimedia Security Technologies for Digital Rights Management*, Boston, Elsevier, 2006
[48] D. Zhu, J. Jung, and D. Song, "TaintEraser: Protecting Sensitive Data Leaks Using Application-Level Taint Tracking,"*ACM Special Interest Group on Operating Systems (SIGOPS) Operating System Review*, vol. 45, no. 1, pp. :142-154, 2011

Solving the Problem of the K Parameter in the KNN Classifier Using an Ensemble Learning Approach

Ahmad Basheer Hassanat[1], Mohammad Ali Abbadi[2],
Ghada Awad Altarawneh[3]
[12]IT Department, [3]Accounting department
Mu'tah University
Mu'tah – Karak, Jordan .

Ahmad Ali Alhasanat
College of Business Administration & Economics
Al-Hussein Bin Talal University,
Maan, Jordan

***Abstract*— This paper presents a new solution for choosing the K parameter in the k-nearest neighbor (KNN) algorithm, the solution depending on the idea of ensemble learning, in which a weak KNN classifier is used each time with a different K, starting from one to the square root of the size of the training set. The results of the weak classifiers are combined using the weighted sum rule. The proposed solution was tested and compared to other solutions using a group of experiments in real life problems. The experimental results show that the proposed classifier outperforms the traditional KNN classifier that uses a different number of neighbors, is competitive with other classifiers, and is a promising classifier with strong potential for a wide range of applications.**

Keywords- KNN; supervised learning; machine learning; ensemble learning; nearest neighbor;

I. INTRODUCTION

The nearest neighbor approach was first introduced by [1] and later studied by [2]. This approach is one of the simplest and oldest methods used for pattern classification. It often yields efficient performance and, in certain cases, its accuracy is greater than state-of the-art classifiers [3] [4].

The KNN classifier categorizes an unlabelled test example using the label of the majority of examples among its k-nearest (most similar) neighbors in the training set. The similarity depends on a specific distance metric, therefore, the performance of the classifier depends significantly on the distance metric used [5].

The KNN classifier is one of the most popular neighborhood classifiers in pattern recognition [6] and [7], because the technique is very simple, and highly efficient in the field of pattern recognition, machine learning, text categorization, data mining, object recognition, etc. [8] and [9]. However, it has limitations, such as memory requirement and time complexity, because it is fully dependent on every example in the training set.

There are two major problems inherited from the design of the KNN [10] and [7]:

1. There is no output trained model to be used; the algorithm has to use all the training examples on each test, therefore its time complexity is linear O(*n*).
2. Its classification performance depends on choosing the optimal number of neighbors (*k*), which is different from one data sample to another.

Many studies have attempted to solve the first problem, dependent on reducing the size of the training set [11], [12], [4], [13] and [14]. Hart proposed a simple local search method called the "Condensed Nearest Neighbor" (CNN) which attempts to minimize the number of stored examples and stores only a subset of the training set to be used for classification later. Their idea is based on removing the similar redundant examples [11].

Gate presented the "Reduced Nearest Neighbor" (RNN) method, which is basically based on the CNN. The aim of the method is to further shrink the CNN stored subset by removing all examples from the subset that do not affect the accuracy of the classifier, i.e. removing them causes no significant error overall [12].

Other studies in the same vein include [15], [16], [17], [18] and [19]. Other works used some hashing techniques to increase classification speed; this includes the work of [20] and [21].

On the other hand, to the best of the authors' knowledge, there has been little work in the literature focuses on the second problem; therefore, the purpose of this study is to solve the second problem of the KNN classifier, by removing the need for using a specific *k* with the classifier.

II. RELATED WORK

Usually, the K parameter in the KNN classifier is chosen empirically. Depending on each problem, different numbers of nearest neighbors are tried, and the parameter with the best performance (accuracy) is chosen to define the classifier.

Choosing the optimal *K* is almost impossible for a variety of problems [22], as the performance of a KNN classifier varies significantly when K is changed as well as the change of distance metric used. However, it is shown in the literature that when the examples are not uniformly distributed, determining the value of K in advance becomes difficult [23].

Guo *et al*. converted the training set to another smaller domain called the "KNN Model". Their model groups each number of similar examples from the data set, based on their

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

similarity to each other. The output model consists of tuples containing the class of the group, the similarity of the most distance point inside the group (local region) to the central data point, in addition to the number of the points of that group (region). There is no need to choose the best k, because the number of points in each group can be seen as an optimal k, i.e. different parameters are used in each group. This work is tested using six data sets, obtaining good results. Their work reduces the size of the training data, and removes the need for choosing the k parameter [10]. However, there is still a need to define other thresholds such as "error tolerant degree" and the minimum number of points allowed in each group.

Song *et al*. presented two approaches – (local informative-KNN (LI-KNN) and global informative-KNN (GI-KNN)) – to solve the problem of the k parameter in the KNN classifier. Their goal was to improve the performance of the KNN. They used a new concept, which they called "*Informativeness*". This was used as a query-based distance metric. Their experiments (based on 10 data sets from the benchmark corpus [24]) showed that their methods were less sensitive to the change of parameters than the conventional KNN classifier [22].

Hamamoto *et al*. used a bootstrap method for nearest neighbor classifier. Their experimental results showed that the nearest neighbor classifier based on the bootstrap samples outperforms the conventional KNN classifiers, mainly when the tested examples are in high dimensions [3].

Yang and Liu argue that the performance of the KNN classifier is relatively stable when choosing a large number of neighbors. They used large values for the *k* parameter such as (30, 45 and 60), and the best results of the classifier were included in their results tables [25] and [26].

Enas and Choi show that the best choice of the *k* parameter was found to be dependent on several factors, namely, the dimension of the sample space, the size of the space, the covariance structure, as well as the sample proportions [27].

The "inverted indexes of neighbors classifier" (IINC) [28], [29] and [30] is one of the best attempts found in the literature to solve the problem. The aim of their work was not intentionally to solve the problem of the k parameters; rather it was designed to increase the accuracy of the classifier. The main idea of the IINC is to use all the neighbors in the training set, rewarding the nearest neighbors, and penalizing the furthest one.

Their algorithm works as follows: the similarity distance of the test point is calculated with all the points in the training set. The distances are sorted in ascending order, keeping track of their classes. The summation of the inverted indexes is then calculated for each class using Eq(1). The probability of each class is then calculated using Eq(2). Obviously, the class with the highest probability is then predicted.

Remark 1*: the previous approach is based on the hypothesis that the influence of the nearest neighbors is larger than those of the furthest distance from the query point* [2], [28], [29] and [30].

The summation of the inverted indexes for class c is:

$$S_c = \sum_{i=1(c)}^{L_c} \frac{1}{i} \quad (1)$$

where L_c is the number of points of class *c*, *i* is the order of the point in the training set after sorting the distances.

The probability of a test point x belongs to a class c can be estimated as:

$$P(x|c) = \frac{S_c}{S} \quad (2)$$

where $S = \sum_{i=1}^{N} \frac{1}{i}$

and N is the number of examples in the training set.

Jirina and Jirina argue that the experimental results based on 24 data sets taken from the benchmark corpus [24], showed that (in most tasks) the IINC outperformed some other well known classifiers such as the traditional KNN, support vector machines, decision trees, artificial neural networks, and naive Bayes classifiers. Therefore there can be an alternative to standard classification methods [28], [29] and [30].

III. THE PROPOSED WORK

There are three problems associated with the reported IINC:

1. It requires all the points in the training data to be used to calculate all the inverted indices; this prevents any attempt to reduce the size of the training set and enforces time consuming.
2. There is bias against the class of the smallest number of points; even if some of those points are around the query point, still the points far away from the query point somehow contribute to increase the probability of the class of the largest number of points. Even if each single contribution of each point get smaller as the points go further, when adding together with large number of points (examples) the contribution become significant.
3. Distances need to be sorted in ascending order to calculate the inverted indices; this take as long a time, at least O(*nlogn*) if quick sort is used; this is worse than the traditional KNN algorithm, which takes a linear time.

We propose to use ensemble learning using the same nearest neighbor rule. Basically, the traditional KNN classifier is used each time with a different K. Starting from k=1 to k = the square root of the training set, each classifier votes for a specific class. Then our multi classifiers system uses majority rule to identify the class, i.e. the class with the highest number of votes (by 1-NN, 3-NN, 5-NN… √n-NN) is chosen.

We choose to have a maximum number of classifiers to be not greater than the square root of the training data set size, because the often used rule of thumb is that k equals the square root of the number of points in the training data set [28], [29] and [30]. Another reason is that more classifiers increases computation time. This complies with what the pilot study

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

suggests, since using this threshold was based on benefit cost, the highest accuracy with the lowest computation time.

The proposed multi classifiers system uses the odd numbers for the k parameter for three reasons: 1) to increase the speed of the algorithm by avoiding the even classifiers; 2) to avoid the chance of two different classes having the same number of votes; and 3) the pilot experiments having the even *k*s show no significant change of the results.

Recalling remark (1), the proposed classifier gives higher weights to the decision of classifiers with the nearest neighbors. The weighted sum rule is used to combine the KNN classifiers. Empirically, we found the best weighting function is using the inverted logarithmic function as in Eq(3). Figure 1 illustrates the function used.

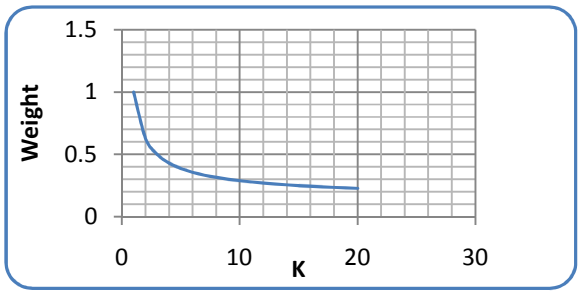


Figure 1. Inverted logarithmic function as weighting function

$$w(k) = \frac{1}{\log_2(1+k)} \tag{3}$$

When a test example is compared with all examples in the training set, using a distance function, an array (A) is created to contain the nearest √n classes, and the weighted sum (WS) rule is defined for each class using:

$$WS_c = \sum_{k=1}^{\sqrt{n}} \sum_{i=1}^{k} \begin{cases} w(i), & A_i = c \\ 0, & otherwise \end{cases}, k = k + 2 \tag{4}$$

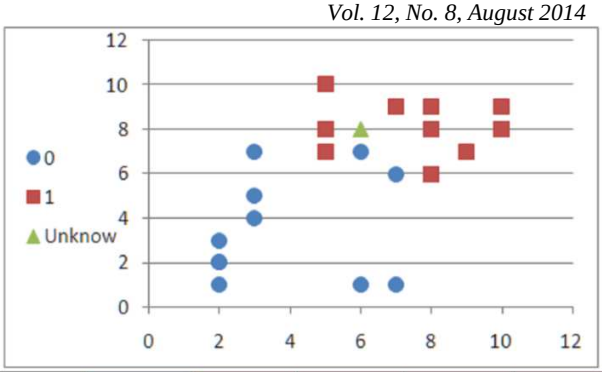
where for each class, we have the outer sum representing the KNN classifier for each odd k, and the inner sum calculates the weights for each classifier.

By applying Eq(4), the highest the votes for a class the highest its WS, and the nearest an example (belonging to a class) to the test example the highest its WS will be. Therefore, the predicted class is the one with the maximum weighted sum:

$$class = \text{argmax}_c \, WS_c \tag{5}$$

To illustrate the proposed classifier, assume that we have 25 points in 2 dimensional feature space belonging to 2 different classes, in addition to one test point (the green triangle) as shown in the upper section of Figure 2.

As shown in Figure 2, the ensemble system uses the 1-NN, 3-NN and 5-NN classifiers using the weighted sum rule to find the class of the unknown point the (green triangle), which in this example is predicted to be class 1 (red square).



class	0	1	1	1	1
order	1	2	3	4	5
weight	1	0.63	0.5	0.43	0.39
k	WS_0	WS_1			
1	1	0			
3	1	1.13			
5	1	1.95			
WS	3	**3.08**			

Result class=1

Figure 2. Simple example showing the proposed classifier

Algorithm 1: *The proposed ensemble KNN classifier*

Input: training data set TD, test example TE

Output: class's index

1. **Array** Distances[n=Size(TD]
2. index=0
3. **For each** example as E in TD {
4. Distances[index]=distanc(E,TE)//*any distance // function*
5. index=index+1
6. }
7. Array minClasses[√n]
8. minClasses = classes (min √n Distances) //*ordered by // distance*
9. **Array** SW[number of classes in TD]// *weight sum for // each class*
10. Initililze SW// *fill with zeros*
11. **for** k=1 **to** √n , k=k+2
12. **for** i=1 **to** k , i=i+1
13. SW[minClasses[i]]=classes[minClasses[i]]+1/Log(1+i,2)
14. **return** argmax(classes)

Based on the time complexity analysis of algorithm 1, we can state the following theorem.

Theorem 1: *Time complexity of the proposed ensemble KNN classifier can be approximated to linear function O(n).*

Proof: Obviously, lines 1 and 2 take O(*1*), lines 3,4 and 5 take O(*n*), n is the size of the training data. Line 7 consumes O(*1*).

Line 8 consumes O(*nlog*√*n*) if we iterate the distance array n times, and insert each element into a binary search tree

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

bounded with size √n, and remove the maximum number when the size of the tree exceeds √n.

Since √n<<n, O(*log*√*n*) can be approximate to a constant *k*, therefore line 8 consumes O(*n k*).

Line 9 consumes O(*1*). Line 10 consumes O(*m*), where m is the number of classes in the training set, which normally is a constant. Thus it can be approximated to O(*1*).

Line 11 consumes O(√*n/2*) because it works only on the odd numbers, The nested loop in line 12 and the line inside (13) consume O(√*n/2**√*n/2*) = O(*n*). And the last line consumes O(*1*).

This makes the total time complexity:

$$2O(1)+3O(n)+O(1)+O(n\ k)+\ 2O(1)+O(\sqrt{n}/2)+2O(n)+O(1)\approx O(n\ k) \quad (6)$$

We can write O(*k*) ≈ O(*1*), therefore:

$$O(n\ k) \approx\ O(n)\ \square$$

The time complexity of the proposed classifier O(*n log* √*n*) ≈ O(*n*) is better than that of the IINC, which is O(*n log n*), because we use only the first √n nearest distances . However, if we worked the naïve version of finding the minimum k distances each time from n elements, it would then cost O(*kn*), since k=√n, time complexity becomes O(*n*√*n*). This function grows even faster than O(*nlogn*).

IV. RESULTS AND DISCUSSION

The proposed classifier is applied and compared to other methods that are found in the literature to solve the problem of the k parameter in the KNN classifier. For the experiments, we chose 28 different data sets to represent real life classification problems, taken from the UCI Machine Learning Repository [24]. Table 1 depicts the data sets used.

TABLE I. DESCRIPTION OF THE DATA SETS USED.

Name	#E	#F	#C	data type	Min	Max
Heart	270	25	2	pos integer	0	564
Balance	625	4	3	pos integer	1	5
Cancer	683	9	2	pos integer	0	9
German	1000	24	2	pos integer	0	184
Liver	345	6	2	pos integer	0	297
Vehicle	846	18	4	pos integer	0	1018
Vote	399	10	2	pos integer	0	2
BCW	699	10	2	pos integer	1	13454352
Haberman	306	3	2	pos integer	0	83
Letter recognition	20000	16	26	pos integer	0	15
Wholesale	440	7	2	pos integer	1	112151
Australian	690	42	2	pos real	0	100001
Glass	214	9	6	pos real	0	75.41
Sonar	208	60	2	pos real	0	1
Wine	178	13	3	pos real	0.13	1680
EEG	14980	14	2	pos real	86.67	715897
Parkinson	1040	27	2	pos real	0	1490
Iris	150	4	3	pos real	0.1	7.9
Diabetes	768	8	2	real & integer	0	846
Monkey1	556	17	2	binary	0	1
Ionosphere	351	34	2	real	-1	1
Phoneme	5404	5	2	real	-1.82	4.38
Segmen	2310	19	7	real	-50	1386.33
Vowel	528	10	11	real	-5.21	5.07
Wave21	5000	21	3	real	-4.2	9.06
Wave40	5000	40	3	real	-3.97	8.82
Banknote	1372	4	2	real	-13.77	17.93
QSAR	1055	41	2	real	-5.256	147

#E: Number of examples. #F: Number of features. #C: Number of classes.

Each data set is divided into two data sets– one for training and the other for testing. 30% of the data set is used for testing, and the rest of the data is for training. Ten types of classifiers have been designed to compare their performances with the proposed classifier; these are 1-NN, 3-NN, 5-NN, 7-NN, 9-NN, √n –NN, 30-NN, 45-NN, 60-NN, and the IINC. These include the traditional KNN classifier using small, medium and large number of neighbors, in addition to the IINC classifier, which arguably bests state-of-the-art classifiers [28], [29] and [30].

Each classifier is used to classify the test samples using Manhattan distance. The 30% of data which were used as a test sample are chosen randomly and each experiment on each data set is repeated 10 times to obtain random examples for testing and training. Table 2 shows the results of the experiments. The accuracy of each classifier on each *normalized* data set is the average of 10 runs.

As can be seen from the results, there is no optimal k, as there is no specific number of neighbors that are suitable for all data sets to be used with the nearest neighbor rule. Each data set favors a specific number (k) of neighbors. This note justifies the proposed method, which attempts to use the power of each classifier, and employs it to enhance the overall performance of the proposed method.

According to the experiments, the using k = √n did not yield excellent results compared to other methods, so using k = √n as a rule of thumb is not a good choice for the KNN classifier. In addition to the use of a large number of neighbors such as k= 30, 45 and 60, does not help in increasing the accuracy of the KNN classifier as argued by [25] and [26]. They argued that the performance of the KNN becomes more stable when using large k. Perhaps that is because their reported results were based on text categorization data sets, while none of the above-mentioned data sets is related to the text categorization problem. Therefore, we cannot generalize their note to other data sets and classification problems.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

TABLE II. THE RESULTS OF THE PROPOSED CLASSIFIER COMPARED TO OTHER CLASSIFIERS– ACCURACIES ARE THE AVERAGE OF 10 RUNS

Data set	1-NN	3-NN	5-NN	7-NN	9-NN	√n -NN	30-NN	45-NN	60-NN	IINC	Proposed
Australian	0.8	0.86	**0.87**	0.86	0.86	0.86	0.85	0.86	0.86	**0.87**	**0.87**
Balance	0.8	0.8	0.83	0.85	0.86	0.88	**0.88**	**0.88**	**0.88**	**0.88**	0.86
Banknote	**1**	**1**	**1**	**1**	**1**	0.98	0.98	0.97	0.96	**1**	**1**
BCW	**0.97**	**0.97**	0.96	0.96	0.96	0.95	0.96	0.96	0.95	0.95	0.96
Cancer	0.96	**0.97**	**0.97**	0.96	0.97	0.96	0.96	0.96	0.95	0.95	0.96
Diabetes	0.69	0.72	0.73	0.74	0.75	**0.76**	**0.76**	**0.76**	0.75	0.74	0.74
EEG	0.84	**0.85**	0.84	0.84	0.84	0.76	0.81	0.8	0.79	0.84	0.83
German	0.69	0.7	0.72	0.73	0.73	0.74	0.71	0.71	0.7	**0.74**	**0.74**
Glass	0.65	0.66	0.66	0.65	0.64	0.64	0.6	0.53	0.42	**0.68**	0.67
Haberman	0.69	0.69	0.72	0.73	0.74	**0.76**	0.74	0.72	0.72	0.75	0.72
Heart	0.76	0.78	0.78	0.79	0.8	0.81	**0.83**	0.82	**0.83**	0.79	0.79
Ionosphere	**0.9**	0.89	0.89	0.89	0.87	0.85	0.84	0.78	0.75	0.85	0.89
Iris	0.94	**0.96**	**0.96**	**0.96**	**0.96**	**0.96**	0.95	0.95	0.88	**0.96**	**0.96**
Letter-recognition	**0.95**	**0.95**	**0.95**	0.94	0.94	0.82	0.91	0.9	0.87	**0.95**	0.94
Liver	0.61	0.63	0.65	**0.66**	**0.66**	**0.66**	0.64	0.64	0.64	0.64	0.64
Monkey1	0.79	0.84	0.91	0.95	**0.96**	0.92	0.91	0.87	0.9	0.92	0.94
Parkinson	0.89	0.91	0.92	0.92	0.92	0.9	0.89	0.9	0.88	**0.93**	**0.93**
Phoneme	**0.89**	0.88	0.87	0.87	0.86	0.83	0.84	0.83	0.83	0.87	0.87
QSAR	0.82	0.84	0.86	0.85	0.85	0.84	0.82	0.81	0.81	**0.86**	**0.86**
Segmen	**0.97**	**0.97**	0.96	0.96	0.96	0.91	0.92	0.91	0.89	0.96	0.96
Sonar	**0.87**	0.83	0.81	0.78	0.75	0.73	0.75	0.72	0.71	0.86	0.85
Vehicle	0.67	0.67	0.66	0.66	**0.68**	0.66	0.65	0.62	0.59	0.67	0.67
Vote	0.91	0.93	0.93	**0.94**	**0.94**	0.93	0.9	0.89	0.89	0.93	0.93
Vowel	**0.98**	0.94	0.87	0.78	0.69	0.53	0.46	0.43	0.38	0.96	0.94
Waveform21	0.76	0.79	0.81	0.82	0.83	**0.85**	0.84	0.84	**0.85**	0.83	0.84
Waveform40	0.71	0.75	0.78	0.79	0.8	**0.84**	0.83	0.83	**0.84**	0.82	0.83
Wholesale	0.86	0.9	**0.91**	**0.91**	**0.91**	0.9	0.89	0.89	0.89	0.9	**0.91**
Wine	**0.97**	0.95	0.96	0.96	0.96	0.96	0.96	0.96	0.94	0.97	0.96
Average	*0.83*	*0.84*	*0.85*	*0.85*	*0.85*	*0.83*	*0.82*	*0.81*	*0.8*	***0.86***	***0.86***

On the other hand, the performance of both the proposed method and the IINC is better than the other classifiers in general. Both methods do not ask for a specific k. The good performance of the IINC is justified by the use of all the neighbors, and the good performance of the proposed method is justified by the use of ensemble learning, which makes use of weak classifiers to generate a stronger one.

It can be noted from the results that the proposed method outperformed all classifiers in 8 data sets, and even when it is behind other classifiers the difference is not more than 0.02 from the best performance. The performance of the IINC is slightly better than the proposed method, as it outperformed all classifiers in 9 data sets. However, both methods have almost the same performance in general.

It is well established in the literature [31] and according to the 'no free lunch' theorem [32], there is no optimal classifier that works perfectly for every class of problems, as the performance of the classifier depends mainly on the problem and the data used.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Our method has yet another feature, which is the linear time complexity, compared to logarithmic linear time of the IINC, which needs to sort the distances to start calculating the inverted indexes. Moreover, the need for all examples in the training set prevents the IINC from speeding up using some methods such as the CNN and RNN. On the other hand, the proposed method can benefit from such methods, because it uses only the square root of the nearest neighbors.

V. CONCLUSION AND FUTURE WORK

This work proposes a new classifier based on the KNN classifier, which solves the problem of choosing the number of neighbors that participate in the final decision using the majority rule of the nearest neighbor approach. The proposed method makes use of the ensemble learning approach, where the traditional KNN is used with a different number of neighbors each time.

The experimental results using a variety of data sets of real life problems have demonstrated the superiority of the proposed method over the tradition KNN using variety of k neighbors. In addition, the proposed method was found competitive to other classifiers such as the IINC classifier. Moreover, we have shown that the speed of the proposed method (linear time) was found to be better than that of the IINC which is logarithmic linear time.

There is room for enhancing the complexity time of the proposed method using KD-trees [33] or other hashing techniques [20] and [21]. Such efforts are best left to be done in the future.

ACKNOWLEDGMENT

All the data sets used in this paper were taken from the UCI Irvine Machine Learning Repository [24], therefore the authors would like to thank and acknowledge the people behind this great corpus. Also the authors would like to thank the anonymous reviewers of this paper.

REFERENCES

[1] E. Fix and J. Hodges, "Discriminatory Analysis: Nonparametric Discrimination: Consistency Properties," 4, 1951.

[2] T. M. Cover and P. E. Hart, "Nearest Neighbor Pattern Classification," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 21-27, 1967.

[3] Y. Hamamoto, S. Uchimura, and S. Tomita, "A Bootstrap Technique for Nearest Neighbor Classifier Design," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 19, no. 1, pp. 73-79, 1997.

[4] E. Alpaydin, "Voting Over Multiple Condensed Nearest Neoghbors," *Artificial Intelligence Review*, vol. 11, pp. 115-132, 1997.

[5] K. Q. Weinberger and L. K. Saul, "Distance Metric Learning for Large Margin Nearest Neighbor Classification," *Journal of Machine Learning Research*, vol. 10, pp. 207-244, 2009.

[6] A. Kataria and M. D. Singh, "A Review of Data Classification Using K-Nearest Neighbour Algorithm," *International Journal of Emerging Technology and Advanced Engineering*, vol. 3, no. 6, pp. 354-360, 2013.

[7] N. Bhatia and A. Vandana, "Survey of Nearest Neighbor Techniques," *(IJCSIS) International Journal of Computer Science and Information Security*, vol. 8, no. 2, pp. 302-305, 2010.

[8] A. B. A. Hassant, "Visual Speech Recognition," in *Speech Technologies*, I. Ipsic, Ed. Rijeka: InTech - Open Access Publisher, 2011, vol. 2, ch. 14.

[9] A. B. A. Hassanat, "Visual Passwords Using Automatic Lip Reading," *International Journal of Sciences: Basic and Applied Research (IJSBAR)*, vol. 13, no. 1, pp. 218-231, 2014.

[10] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN Model-Based Approach in Classification," *Lecture Notes in Computer Science*, vol. 2888, pp. 986-996, 2003.

[11] P. Hart, "The Condensed Nearest Neighbour Rule," *IEEE Transactions on Information Theory*, vol. 14, pp. 515-516, 1968.

[12] G. Gates, "The Reduced Nearest Neighbour Rule," *IEEE Transactions on Information Theory*, vol. 18, pp. 431-433, 1972.

[13] M. Kubat and M. Jr, "Voting Nearest-Neighbour Subclassifiers," in *Proceedings of the 17th International Conference on Machine Learning, ICML-2000*, Stanford, CA, 2000, pp. 503-510.

[14] D. R. Wilson and T. R. Martinez, "Reduction Techniques for Exemplar-Based Learning Algorithms," *Machine learning*, vol. 38, no. 3, pp. 257-286, 2000.

[15] Y. Zeng, Y. Yang, and L. Zhou, "Pseudo Nearest Neighbor Rule for Pattern Recognition," *Expert Systems with Applications*, vol. 36, pp. 3587-3595, 2009.

[16] H. Parvin, H. Alizadeh, and B. Minaei, "A Modification on K-Nearest Neighbor Classifier," *Global Journal of Computer Science and Technology*, vol. 10, no. 14, pp. 37-41, 2010.

[17] Z. Yong, "An Improved kNN Text Classification Algorithm based on Clustering," *Journal of Computers*, vol. 4, no. 3, 2009.

[18] Q.-B. Gao and Z.-Z. Wang, "Center-based nearest neighbor classifier," *Pattern Recognition*, vol. 40, pp. 346-349, 2007.

[19] Z. Yong, L. Youwen, and X. Shixiong, "An Improved KNN Text Classification Algorithm Based on Clustering," *JOURNAL OF COMPUTERS*, vol. 4, no. 3, pp. 230-237, 2009.

[20] P. Indyk and R. Motwani, "Approximate nearest neighbor: towards removing the curse of dimensionality," in *Proc. 30th Annu. ACM Symp. Comput. Geometry*, 1998, p. 604–613.

[21] A. Andoni and P. Indyk, "Near-Optimal Hashing Algorithms for Approximate Nearest Neighbor in High Dimensions," *COMMUNICATIONS OF THE ACM*, vol. 51, no. 1, pp. 117-122, 2008.

[22] Y. Song, J. Huang, D. Zhou, H. Zha, and C. L. Giles, "Iknn: Informative k-nearest neighbor pattern classification," *in Proceedings of the 11th European conference on Principles and Practice of Knowledge Discovery in Databases*, Berlin, 2007, pp. 248-264.

[23] M. Latourrette, "Toward an explanatory similarity measure for nearest-neighbor classification," in *Proceedings of the 11th European Conference on Machine Learning*, London, 2000, pp. 238-245.

[24] K. Bache and M. Lichman. (2013) UCI Machine Learning Repository. [Online]. http://archive.ics.uci.edu/ml

[25] Y. Yang, "An evaluation of statistical approaches to text categorization," *Information Retrieval*, vol. 1, pp. 69-90, 1999.

[26] Y. Yang and X. Liu, "A re-examination of text categorization methods," in *Proceedings of SIGIR-99, 22nd ACM International Conference on Research and Development in Information Retrieval*, Berkeley, 1999, pp. 42-49.

[27] G. G. Enas and S. C. Choi, "Choice of the smoothing parameter and efficiency of k-nearest neighbor classification," *Computers &*

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Mathematics with Applications, vol. 12, no. 2, pp. 235-244, 1986.

[28] M. Jirina and M. J. Jirina, "Classifier Based on Inverted Indexes of Neighbors," Institute of Computer Science, Technical Report No. V-1034, 2008.

[29] M. Jirina and M. J. Jirina, "Using Singularity Exponent in Distance Based Classifier," in *Proceedings of the 10th International Conference on Intelligent Systems Design and Applications (ISDA2010)*, Cairo, 2010, pp. 220-224.

[30] M. Jirina and M. J. Jirina, "Classifiers Based on Inverted Distances," in *New Fundamental Technologies in Data Mining*, K. Funatsu, Ed. InTech, 2011, vol. 1, ch. 19, pp. 369-387.

[31] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Wiley, 2001.

[32] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE Trans. Evol. Comput.*, vol. 1, p. 67–82, 1997.

[33] J. L. Bentley, "K-d trees for semidynamic point sets,"*In Proceedings of Sixth Annual Symposium on Computational Geometry*, 1990, pp. 187-197.

AUTHORS PROFILE

Ahmad B. A Hassanat was born and grew up in Jordan, received his Ph.D. in Computer Science from the University of Buckingham at Buckingham, UK in 2010, and B.S. and M.S. degrees in Computer Science from Mutah University/Jordan and Al al-Bayt University/Jordan in 1995 and 2004, respectively. He has been a faculty member of Information Technology department at Mutah University since 2010. His main interests include computer vision, Machine learning and pattern recognition.

Mohammad Ali Abbadi received his Ph.D and M.S. in computer science from George Washington University, USA, in 2000, and 1996 respectivly, and B.S. in computer science in 1990 from Mutah University. He has been a faculty member of Information Technology department at Mutah University since 2000. His research interests include Data Compression, Multimedia Databases & Digital Libraries, Audio/Image/Video Processing, Operating Systems, Fault Tolerance and Watermarking.

Ahmad Ali Alhasanat received his M.S. degree in Management Information Systems from Al-Balqa' Applied University/Jordan in 2014, and B.S. degree in Computer Science from Mutah University/Jordan in 2003. He has been a computer lab supervisor in College of Business Administration & Economics at Al-Hussein Bin Talal University since 2004. His main interests include management information systems and artficial intelligence.

Ghada Awad Altarawneh received her Ph.D. in Accounting from the University of Buckingham at Buckingham, UK in 2011, and B.S. and M.S. degrees in Accounting from Mutah University/Jordan and Al al-Bayt University/Jordan in 2002 and 2005, respectively. She has been a faculty member of Accounting department at Mutah University since 2011. Her main interests include Accounting information systems, and using artificial intelligence in accounting systems.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Proposing a New Hybrid Approach in Movie Recommender System

Monireh Amini
Department of Computer Engineering,
Zanjan Branch, Islamic Azad University,
Zanjan, Iran

Mahdi Nasiri
Computer Engineering Department
Iran University of Science and Technology
Tehran, Iran

Mahdi Afzali
Department of Computer Engineering,
Zanjan Branch, Islamic Azad University,
Zanjan, Iran

***Abstract*—Due to the unprecedented growth of information, goods and services, a lot of application programs have been created in recent years to help the selection of goods and services to customers. One of the most important application programs are recommender systems that used to things as proposed movies, books, web pages, and E-Business, etc. Most of recommender systems using Collaborative filtering (CF) and or content based filtering (CBF) to provide suggestion for users. In this paper a new approach examined for better assessing the interests of customers. With the understanding of customer behavior, appropriate offer will be provided to customers. In fact, by using a new hybrid approach, weakness of content based filtering and Collaborative filtering methods as much as possible to will be resolve. The results of this paper can be used to keep and attract customers in the institutions or stores that have fixed customers. In this paper, first we reviewing the recommender systems and investigating types of filtering. Then a new hybrid approach by using CF and CBF methods is presented in a movie recommender system. Results are evaluated on movielens valid data, that the results show improvement in the movie recommender system.**

Keywords-Hybrid Recommender Systems; Collaborative filtering; Content-based filtering; Hybrid filtering; Spiking Neural Network (SNN) ; Naive Bayes; E-Business

I. INTRODUCTION

Recommender Systems have been provided different strategies to buy and how to spend leisure time with the purpose of personal recommendations, quality and affordable. Recommender System use of statistical and knowledge discovery techniques to solve the interaction with the target customers to provide products recommended [1]. A common scenario for modern recommendation systems is a Web application which with a user interacts [2]. A user model is the core of a recommender system, the approach which information is obtained depends on the particular recommendation technique [3]. User preferences can, for instance, be acquired implicitly by monitoring user behavior, but the recommender system might also explicitly ask the visitor about his or her preferences [4]. Recommender system is composed of three main parts: 1) Product information, 2) the exchange of information that at the start of the recommendation process, users can interact with the system, 3) an algorithm that combines product information and user and provides a recommendation [5].

A. Collaborative filtering (CF)

CF is a method of making automatic predictions about the interests of a user by collecting taste information from many users [6]. These methods have been developed for offering intern shop products or audiovisual materials, which have problems with the Straggly rating data [7]. CF methods are divided into two main types: 1) neighborhood or memory-based, 2) model-based [8, 9]. Model-based algorithms for predicting rates, used a learning model [3]. Model-based methods, the utility function is not based on some ad hoc heuristic rules, but based on learned model from the original data, using statistical learning techniques to calculate [8]. Model-based method by analyzing the data, clustered them in the estimated models and using methods such as Bayesian models, neural networks or latent semantic analysis. These methods had accurate prediction than memory–based method, but these are required huge initial investment for estimating models and make accurate recommendations [10]. In neighborhood based (heuristic-based) CF, the user-item ratings stored in the system are directly used to predict ratings for new items. This can be done in two ways known as user-based or item-based recommendation [9]. While the user-based method in prediction rating relies on the users that have same opinion, item-based method relies on the scores is given to similar items [11] and assumes that the user is most likely to buy the same items that are already purchased [12].

B. Content-Based filtering (CBF)

In content-based filtering systems, a user profile represents the content descriptions of items in which that user has previously expressed interest. The content descriptions of items are represented by a set of features or attributes which characterize that item [13]. Content-based information describes the actual data [14].

C. Knowledge-based filtering

Knowledge-based recommendation systems suggest products based on inferences about a user's needs and preferences. These systems have no start-up problems and do not require user ratings. However, knowledge acquisition is very difficult [15]. This method is used deep knowledge about the domain items to determine recommendations [16]. There are two well-known methods to knowledge-based recommendation: case-based recommendation and constraint-based recommendation [17, 4].

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

D. Hybrid filtering

Hybrid recommender systems are technical approaches that combine several algorithm implementations or recommendation component [4]. Hybrid recommender system combined two or more recommendation techniques to gain better performance with fewer of the drawbacks of any individual one. For example, the Recommendz System is combination of collaborative, content-based and knowledge-based filtering [15]. Different hybrid strategies are [4]:

- Weighted: A weighted hybridization strategy combines the recommendations of two or more recommendation systems by computing weighted sums of their scores.
- Switching: In this method, according to the current, conditions system selected one of the recommended methods. Switching hybrids require an oracle that decides which recommender should be used in a specific situation, depending on the user profile and/or the quality of recommendation results.
- Mixed: A mixed hybridization strategy combines the results of different recommender systems at the level of the user interface, in which results from different techniques are presented together.
- Feature Combination: A feature combination hybrid is a monolithic recommendation component that uses a diverse range of input data.
- Feature Augmentation: Feature augmentation is another monolithic hybridization design that may be used to integrate several recommendation algorithms. This hybrid does not simply combine and preprocess several types of input, but rather applies more complex transformation steps. In fact, the output of a contributing recommender system augments the feature space of the actual recommender by preprocessing its knowledge sources.
- Cascade: Cascade hybrids are based on a sequenced order of techniques, in which each succeeding recommender only refines the recommendations of its predecessor. The recommendation list of the successor technique is thus restricted to items that were also recommended by the preceding technique.
- Meta-level: In a meta-level hybridization design, one recommender builds a model that is exploited by the principal recommender to make recommendations.

Related work is presented in section 2, in section 3 proposed hybrid approach is expressed. In section 4, initial settings, data set, and experiment results are described of experiment. Section 5 discusses the results and conclusions of the paper.

II. Related Work

Recommender systems have become an important research area since the appearance of the first papers on CF in the let of 20 century. This research mainly relied on a user's rating [18]. Examples: recommending books, and other products at Amazon.com [19], movies by Movie Lens [20], and news at VERSIFI Technologies (formerly AdaptiveInfo.com [21]) [18].

The Group Lens project started in 1992 and completed a pilot study at two sites to establish the feasibility of using CF for Usenet news [22]. In 2005, a hybrid Recommender System based on Multi-Layer Perceptron Neural Network with CF and CBF approach is presented in [23]. CF Hybrid Approach using Neural Network was proposed in 2006, which has been used in dataset movielens [24].

A study was conducted with the name of the MoviExplain. This movie recommender system consists of four main subsystems which include: 1) a web crawler, 2) the database profiles, 3) a Recommendation Engine and 4) the website that offers a list of videos to users [25]. The movie recommender system called E-MRS was created based on emotions. The architecture of E-MRS is designed to recommend movies to users based on their emotions. The recommendation algorithm is in fact a cascade hybrid of two techniques: a CF and a content-based recommendation [15].

Web-based movie recommender system was presented in 2007, which used of the three techniques; Demographic filtering (DF), CBF, and CF. The system recommend the 20 movie at any time [26]. DF uses descriptions of people to learn the relationship between a particular item and the type of people who like it [27]. In 2010 a hybrid recommender system using Naive Bayes and item-based CF has been suggested [28]. The hybrid recommender system based on Neural Network is presented by combining CF and CBF [29]. In 2011 recommender algorithm as Android application with additional functions are implemented by combining with existing web services and API is able to display film the cinema scene with details of the information movies for the user[6]. Also hybrid recommender system by combining predictions using CF based on neighborhood, DF and CBF has been presented [30]. In 2013 ORBIT recommender system combining CF and CBF methods have been proposed [31].

III. The Proposed Hybrid Approach

One of the challenges in recommender systems is increase prediction accuracy and precision. In this section, a proposed approach in the hybrid recommender systems will be presented to increasing improvements in forecasting. Hybrid Recommender Systems can use several filtering methods for prediction. By combining several recommendation methods, can decrease the weaknesses of systems that use only same recommendation method. By using combination of several filtering advantage, prediction accuracy and precision is increased. Thus making hybrid recommender systems, which combine the advantage of different filtering to overcome some of the weaknesses and problems, has become purpose of recent research. Various hybrid methods for hybrid recommender systems can be applied, such as CF hybrid approach with CBF, or CF and KB, or DF with CBF.

The proposed approach of this paper, are combination of CBF and CF methods, that proposed approach and the combined methods of CF and CBF is described in the following. The different types of data that will be applied to

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

recommender systems include content-based information and collaborative-based information. Content-based information is involving explanation and feature about items. In content-based recommendation, purposes are patterns between the target user and the content. Collaborative information is including other user comments in data, for example, individual ratings of movies. Some recommender systems that only using from collaborative information to find the correlation between system users and user target. This systems failure, because they only rating correlation between users with same features. The problem is that when the database contains thousands of samples, the target user has rated only a few of them.

Proposed hybrid approach in the paper, used of collaborative information and content-based information. The proposed approach combines of both CF and CBF methods that utilizes a monolithic hybridization design, is given in Fig. 1. In Monolithic hybridization design, hybridization is thus, achieved by a change in algorithm behavior to exploit different types of input data. Monolithic combination includes the aspects of several methods of filtering in an algorithm implementation. Several recommenders contribute because the hybrid uses additional input data that are specific to another recommendation algorithm. Or the input data are augmented by one technique and factually exploited by the other. Monolithic hybridization design include: feature combination strategy and feature augmentation strategy. In the proposed approach, we use of feature combination strategy in the monolithic hybridization design. Feature combination by uses of diverse range of input data. The Proposed hybrid approach involves two steps:

A. *First step: the samples clustering*

The items content features and Collaborative features are used to identify similar models. In ordering to, we apply K-Means clustering algorithm. How to determine the proper number of clusters and clustering results explained in the next section.

B. *Second step: model structure*

After the clustering step, we use the classification models. Data mining techniques involve: predict methods and describe methods. In prediction methods, used of some features to predict the value of a specify feature. In this paper, we use user collaborative features and movie content features to predict the rating of the movie. According to research data, the most appropriate model should be chosen, because each model is more suitable for specific data. Our goal is to predict the rating of the movie by classification model. Models include four classifiers models that are Spiking Neural Network (SNN), Multi-Layer Perceptron Neural Network (MLP), Decision Tree, and Naive Bayes. In the proposed approach, we used method predictions and supervised.

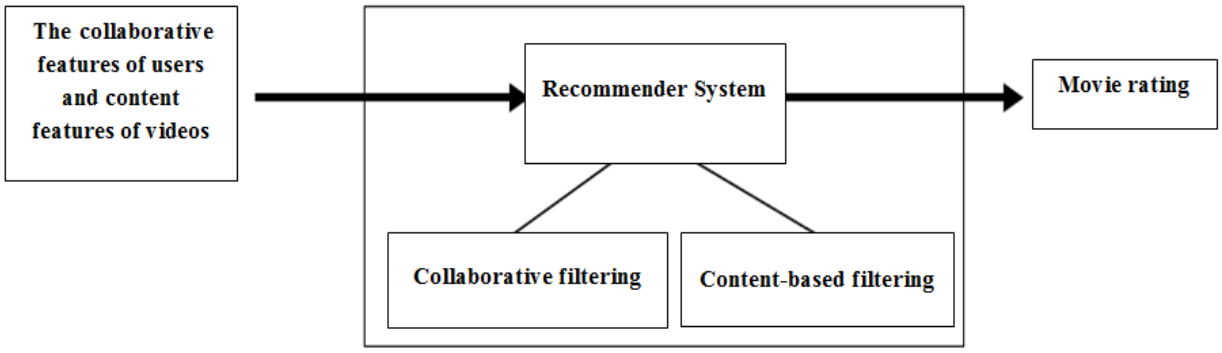


Figure 1. The proposed hybrid approach by using monolithic hybridization design

IV. Experimental Result and Evaluation

A. *Movielens dataset*

The dataset selected for this study is a standard datasets Movielens. This dataset has been established by Lens research group from the University of Minnesota. Information and Movielens dataset accessed of the related websites[1]. This dataset is comprised of three types of datasets. The first dataset used in this paper is the dataset consists of 100,000 ratings for 1682 movies by 943 users. This dataset has been collected from 1997 to 1998, which includes tree data file is as follows:

- File rankings: ranking the file involving the user id, video id, rating and timestamp. In this dataset, each user has rated at least 20 movies.
- User file: this file contains the user id and user's collaborative information, such as gender, age, job and zip code is. All information is provided in partnership with users. So there is no error in it.
- Video files: video files that contain the movie id, movie title, release date and type of film (19 genres).

In this paper, first we extract the dataset, then using a dataset of users, videos, rating, we created a central data

[1] http://www.Movie Lens.org, http://grouplens.org/datasets/movielens/

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

repository. The most important and most common type is the data matrix. The most important and most common type of data is the matrix data. This dataset can be represented by a matrix m * n, so that m rows (each row belongs to one record) and n columns (each column belongs to a feature) has said. For analysis of data we used the data matrix.

B. *Initial settings*

According to each model, determine basic parameters and preprocessing data requirements must be applied to the data, until results can be properly assessed. In this section, the basic parameters of each model and the experiments are given. All preprocessing apply on a central data repository, the most important preprocessing operations to be expressed in continue. There are 19 types of genres movies in the movielens dataset. In feature subset selection, we select 17 important style and we unseen user's zip code. We draw features correlation matrix that if the features are correlated, one of them is removed.

In the experiments, is used from user collaborative features , such as user id, age, gender, job, and also is used from movies features, such as types of movies style including action, Adventure, Children's, comedy, crime, documentary, drama, fantasy, Film-Noir, horror, musical, mystery, romance, Sci-Fi, Thriller, War, Western and movie release date, movie id. After the data preparation step, first we are clustering data by using the k-means clustering method. To select an appropriate number of clusters, with apply different number of clusters in the multi-layer perceptron neural network, we have examined accuracy and precision. The results in Fig. 2 indicate that by determination of 10 clusters, the accuracy and precision of the calculations in the model increased. As a result, the data is divided into 10 clusters, and then evaluates the results by classified models.

By using the techniques of random sampling, we randomly selected 5,000 records with the number of samples, evaluated each four classifiers model. The input data must be encoded for the Spiking Neural network. By using conversion functions, the input data can be converted into spike patterns. Thus, the input data codes in the range of 0 to 6 time unit. And also output data codes in the time range 10 to 16. We applied back-propagation learning techniques for Spiking Neural Networks [32]. Also in the input layer of the Spiking Neural Network 23 neurons, 11 neurons in the middle layer, and there are 4 neurons in outer layer.

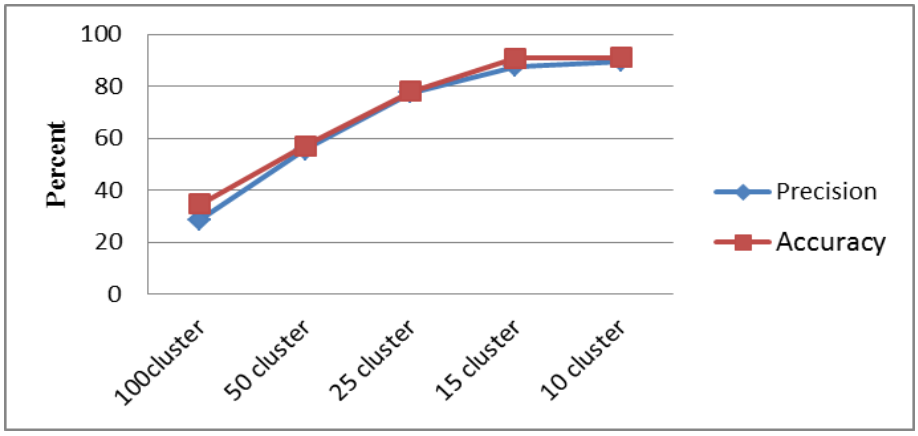


Figure 2. Compares clustering results with the accuracy and precision criterias

C. *The results*

In this paper proposed a hybrid approach using Collaborative filtering content-based filtering. In proposed approach had used of monolithic hybridization design, that this design are combination several knowledge source, include rating information of movie, user's information, movie information. For predicted movie rating, used four classifier models. The models include: Multi-Layer Perceptron Neural Network (MLP), Spiking Neural Network (SNN), Naive Bayes, Decision tree.

Fig. 3 shows the results of the models in terms of classification error. According to Fig. 3, the classification error in Spiking the neural network 0.52% and for the decision tree 0.07%, These models are have more classification error than the other two models. Fig. 4 show results of the three models with accuracy and precision criteria and Recall. According to Fig. 4 and table 1, results predict that the Naïve Bayes model with precision 99.83% and Classification error 0.2% and also MLP with precision 95.55% and Classification error 4.13%.

In this section, the experiments studied on the movielens data base. Also we used 70% of the dataset for training and the 30% for testing. It should be noted that if training data is larger, predicted will be better and test dataset is larger, the error estimation will be accurate.

TABLE I. RESULTS OF THE MODELS

The models	Precision	Recall	Accuracy	Classification error
Naive Bayes	99.83	99.71	99.8	0.2
MLP	95.55	95.51	95.87	4.13
Decision tree	85.8	90	92.61	7.39

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

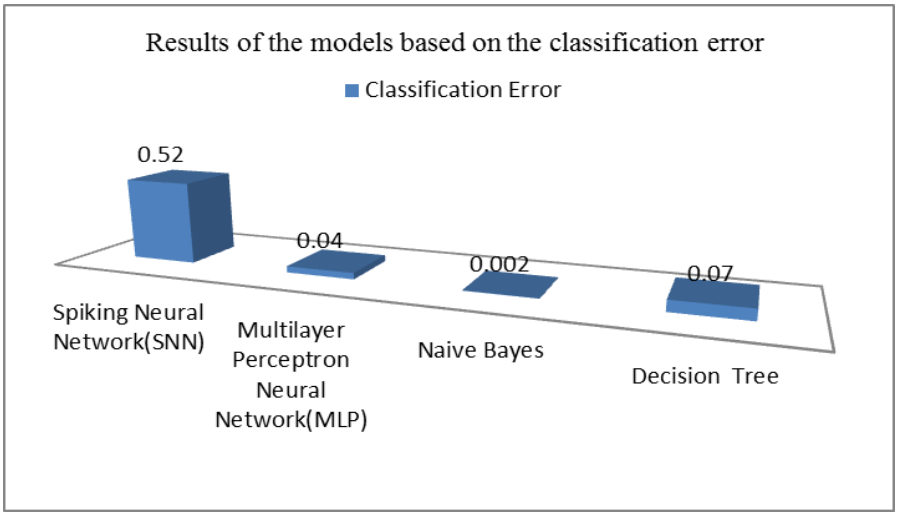


Figure 3. Results of the models based on the classification error

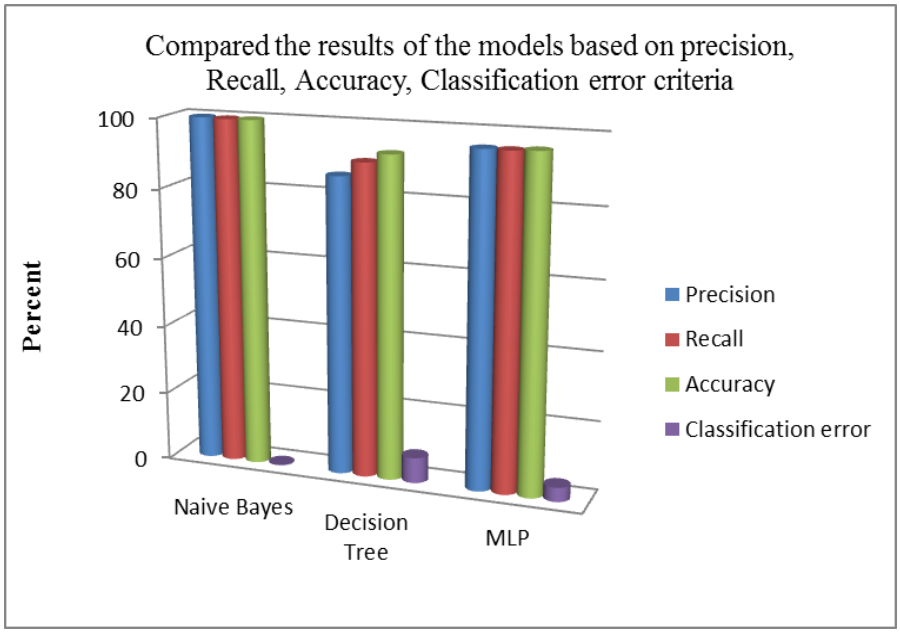


Figure 4. Results of the models with different criteria

V. Discussion and Conclusions

In this paper a new hybrid approach presented in movie recommender system. In the proposed approach, we used from CF and CBF methods. In proposed approach used monolithic hybridization design, that is combines an implementation of different input data. In fact, several knowledge sources compound with user-collaborative information and movie content information. Using the knowledge resources we create a central data repository contains 100,000 records and 24 features. After preprocessing and clustering, predict movie rating by using classifier models. Models include: Spiking Neural Network (SNN), Multi-layer Perceptron Neural Network (MLP), Decision tree, Naive Bayes.

In these experiments, data preprocessing and setting of basic parameters of each model is the most important part of test, that several preprocessing techniques were applied to achieve the optimal results. Create a central data repository is one of the most important steps in data preparation, central data repository including information about the user and movies, and rating data.

Results are indicated improvement in hybrid recommender system and high accuracy and precision in models. According to the model results Naive Bayes is the highest prediction accuracy than other models. Spiking Neural Network (SNN) model used first in recommender systems, than proposed in other fields. The hybridization design in hybrid approach with other designs possible, such

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

as parallel and linear design. These designs and other filtering methods proposed in future works.

REFERENCES

[1] Z. Zhang and S. Qian, "The Research of E-commerce Recommendation System Based on Collaborative Filtering Technology," Advances in CSIE, Vol. 1, AISC 168, Springer-Verlag Berlin Heidelberg, pp. 507–512, 2012.

[2] M. J. pazzani and D. billsus, "content-based recommendation system," The Adaptive Web, LNCS 4321, Springer-Verlag Berlin Heidelberg, pp. 325–341, 2007.

[3] A. Felfering, G. Friedrich and L. Schmidt-Thieme, "Guest Editors' Introduction: Recommender Systems," IEEE Intelligent systems, Volume 22 Issue 3, pp. 18–21, doi: 10.1109/MIS.2007.52 , May 2007.

[4] D. Jannach, M. Zanker, A. Felfernig and G. Friedrich, "Recommender systems : an introduction," Printed in the United States of America, Cambridge University Press, 2011.

[5] W. Hill, L. Stead, M. Rosenstein, and G. Furnas, "Recommending and evaluating choices in a virtual community of use," In Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems, doi: 10.1145/223904.223929, Pages 194-201, New York, USA, 1995.

[6] S.K. Ko, S.M. Choi, H.S. Eom, J.W. Cha, H. Cho, L. Kim, and Y.S. Han, "A Smart Movie Recommendation System, " Human Interface, Part I, HCII 2011, LNCS 6771, Springer-Verlag Berlin Heidelberg, pp. 558–566, 2011.

[7] T. D. Pessemier, S. Coppens, K. Geebelen, C. Vleugels, S. Bannier, E. Mannens, K. Vanhecke, and L. Martens, "Collaborative recommendations with content-based filters for cultural activities via a scalable event distribution platform," Multimedia Tools and Applications, Volume 58, Issue 1, DOI 10.1007/s11042-010-0715-8, 58:167–213. ,2012.

[8] J.S. Breese, D. Heckerman, and C. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," 14th Conf. Uncertainty in Artificial Intelligence, University of Wisconsin, madison, July 1998.

[9] C. Desrosier, and G. Karypis, "A comprehensive survey of neighborhood-based recommendation methods," in Recommender Systems Handbook, Part 1, DOI: 10.1007/978-0-387-85820-3_4, pp. 107-144, 2011.

[10] R. Ghiyasi, H. Hani Zavarehe, "Evaluation and classification methods for creating personalized recommender systems," The Fourth Iran Conference on Electrical Engineering, Iran, Islamic Azad University Gonabad, 2012.

[11] M. Deshpande, and G. Karypis, "Item-based top-N recommendation algorithms". ACM Transaction on Information Systems, Doi: 10.1145/963770.963776, Volume 22 Issue 1, pp. 143–177, January 2004.

[12] Y. Jiang, J. Shang, and Y. Liu, "Maximizing customer satisfaction through an online recommendation system: A novel associative classification model," Decision Support Systems, DOI: 10.1016/j.dss.2009.06.006, Volume 48 Issue 3, pp 470–479, 2010.

[13] B. Mobasher, "Data Mining for Web Personalization," Center for Web Intelligence, Springer-Verlag Berlin Heidelberg, pp. 90–135, 2007.

[14] W. Kogel, "Faster Training of Neural Networks for Recommender Systems," Thesis, Degree of Master of Science in Computer Science, WORCESTER POLYTECHNIC INSTITUTE, May 2002.

[15] A.T. Ho, and I.L.L. Menezes, and Y. Tagmouti, "E-MRS: Emotionbased movie recommender system," Proceedings of IADIS e-Commerce Conference. USA: University of Washington Bothell, 2006.

[16] R. Burke, "Hybrid Recommender Systems: Survey and Experiments," User Modeling and User-Adapted Interaction, Volume 12, Issue 4, pp 331-370, 2002.

[17] A. Felfernig, and R. Burke, "Constraint-based Recommender Systems: Technologies and Research Issues," 10th Int. Conf. on Electronic Commerce (ICEC) '08 Innsbruck, Austria, DOI: 10.1145/1409540.1409544, ACM New York, USA, 2008.

[18] G. Adomavicius, and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 17, NO. 6, pp 734- 749, 2005.

[19] G. Linden, B. Smith, and J. York, "Amazon.com Recommendations Item-to-Item Collaborative Filtering," IEEE Internet Computing, Published by the IEEE Computer Society, pp. 76-82, 2003.

[20] B.N. Miller, I. Albert, S.K. Lam, J.A. Konstan, and J. Riedl, "MovieLens Unplugged: Experiences with an Occasionally Connected Recommender System," In Proceedings of the Conference on Intelligent User Interfaces, Florida, 2003.

[21] N. J. Belkin, and W.B. Croft, "Information filtering and information retrieval: Two sides of the same coin?," Communications of the ACM, vol. 35, no. 12, pp. 29-38, 1992.

[22] P. Resnick, N. Iacovou, M. Sushak, P. Bergstrom, and J. Riedl, "Group- Lens: An open architecture for collaborative filtering of netnews," In Proceedings of the 1994 Computer Supported Cooperative Work Conference, ACM, New York, 1994.

[23] C. Christakou, A. Stafylopatis, "A Hybrid Movie Recommender System Based on Neural Networks," International Conference on Intelligent Systems Design and Applications (ISDA'05), IEEE , 2005.

[24] C. Vassiliou, D. Stamoulis, and D. Martakos, "A Recommender System Framework combining Neural Networks & Collaborative Filtering," Proceedings of the 5th WSEAS Int. Conf. on Instrumentation, Measurement, Circuits and Systems, pp 285-290, 2006.

[25] P. Symeonidis, A. Nanopoulos, and Y. Manolopoulos, "MoviExplain: A Recommender System with Explanations," Proceedings of the third ACM conference on Recommender systems, 317-320, 2006.

[26] N. T. Nguyen, and M. Rakowski, M. Rusin, and J. Sobecki, and L. C. Jain, "Hybrid Filtering Methods Applied in Web-Based Movie recommendation System," Knowledge-Based Intelligent Information and Engineering Systems, Australia, DOI 10.1007/978-3-540-74819-9_26, Volume 4692, pp 206-213, 2007.

[27] M. Montaner, B. Lopez, and J.P. de la Rosa, "A Taxonomy of Recommender Agents on the Interne," Artificial Intelligence Review 19, pp 285–330, 2003.

[28] M.A. Ghazanfar, and A. Prugel-Bennett, "An Improved Switching Hybrid Recommender System Using Naive Bayes Classifier and Collaborative Filtering," International MultiConference of Engineers and Computer Scientists, ISBN: 978-988-17012-8-2, Vol I, Hong Kong, 2010.

[29] Y. Deng, W. Zhonghai, C. Tang, H. Si, H. Xiong, Z. Chen, "A Hybrid Movie Recommender Based on Ontology and Neural Networks," IEEE/ACM International Conference on Green Computing and Communications & International Conference on Cyber, Physical and Social Computing, China, 2010.

[30] B. Chikhaoui, M. Chiazzaro, and SH. Wang, "An Improved Hybrid Recommender System by Combining Predictions," 25th IEEE International Conference on Advanced Information Networking and Applications Workshops, Biopolis, Singapore, 2011.

[31] D. Pathak, S. Matharia, and C.N.S. Murthy, "ORBIT: HYBRID MOVIE RECOMMENDATION ENGINE," IEEE International Conference on Emerging Trends in Computing, Communication and Nanotechnology, Tirunelveli, 2013.

[32] S. M. Bohte, and J. N. Kok, H. L. Poutre, "Error back propagation in temporally encoded networks of spiking neurons, Neurocomputing," Elsevier Science, Vol.48, pp.17-37, 2002.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Trellis Analysis of Transmission Burst Errors in Viterbi Decoding

Salehe I. Mrutu[1], Anael Sam[2] and Nerey H. Mvungi[3]

[1, 2] *School of Computational and Communication Science and Engineering (CoCSE),*
Nelson Mandela African Institution of Science and Technology (NM-AIST), Arusha, Tanzania
[3] *College of Information and Communication Technologies, University of Dar Es Salaam, Dar Es Salaam (UDSM), Tanzania*

***Abstract*—The Viterbi decoder is the most favorable solution to the problem of decoding codewords from a convolutional encoder. Viterbi decoder performs exceptionally well when a received codewords block contains single or multiple and scattered errors in a received codewords block. However, the formation of burst errors in data transmission due to high transmission speed and the widely varying error conditions of wireless media in fading channel creates decoding challenge for such conditions which result in unbearable amount of residual errors. By using Viterbi decoders' trellis diagrams, this paper analyses the effects of burst errors to the decoder that lead to residual errors and proposes improvement to the encoding and decoding procedures of the existing (2, 1, 2) binary convolutional encoder. The improved version facilitate effectiveness in the decoder (Viterbi algorithm) in decoding burst errors and hence reduction of residual errors in a poor channel. The proposed enhancements improve the decoder's operational performance by 75 percent. However, the proposed modification reduces the encoder's data transmission rate from 1/2 to 1/6.**

Keywords-Locked Convolutional encoder; Bust errors; Residual errors; Non Transmittable Codewords (NTCs); Viterbi Algorithm Decoding

I. INTRODUCTION

Viterbi Algorithm (VA) decoder is named after its founder Andrew J. Viterbi[1],[2]. This algorithm was first proposed in 1967 and further developed by the same author in 1971 as a decoding algorithm for binary convolutional codes transmitted over a noisy digital communication channel [3]. Since then, the VA has found a wide range of applications in both satellite and other mobile communication such as Code Division Multiple Access (CDMA) and Global System Mobile (GSM) digital cellular, dial-up modems, satellite, target tracking, deep space communications, 802.11 wireless LANs to mention a few. Apart from communication applications the VA found a numerous application in speech synthesis, speech recognition, keyword spotting, bioinformatics and computational linguistics. Omura [4] and Forney [5] showed that the VA is a maximum likelihood decoder. The basis of VA is on minimizing error probability by comparing the likelihood of a set of possible state transitions that can occur, and judge which one of the results has the highest probability of occurrence. However, in this regard the application of VA decoder to decode convolutional codes transmitted over unreliable transmission channel with noise is discussed. Fig. 1 describes a digital communication system where, the flow of information stream from a sender machine where source information are passed through the source encoder which converts them into binary data for the channel encoder (convolutional encoder). Binary convolutional encoder processes the supplied data and adds redundant bits which are used by channel decoder (Viterbi decoder) in estimating the original data at the receiver. The codewords are output of convolutional encoder. The ratio of input bits to the output bits of the encoder is called the code rate. For this case each binary bit produces two binary bits from the encoder and hence the code rate is 1/2.

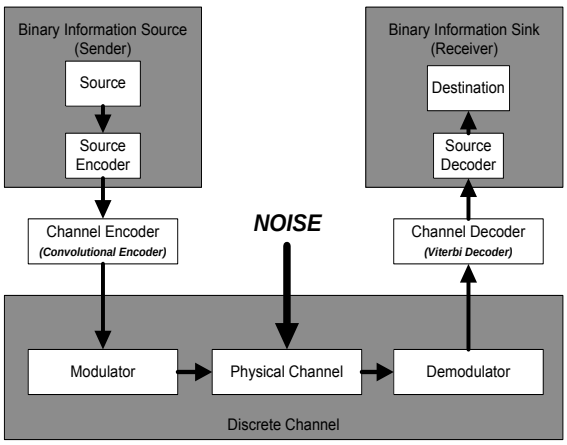


Fig. 1. A block diagram of a digital communication system

From the channel encoder, bit sequences are then modulated to relevant waveforms for transmission through a channel. The characteristics of the channel introduce noise to the waveforms in the channel and distort the waveforms [6] resulting in transmission errors. Transmission errors in digital communication can occur as a single bit error where a bit in a data block is altered or multiple bit errors where many but not successive bits in a data block are altered or burst bit errors where successive bits are altered in a data block [7]. How much distortion can be expected in transmission is a matter of type of channel used and if a wireless channel is used then it is highly dependent on weather condition. Transmission errors cause the received data to be different from the sent one, thus the need for a mechanism of identifying and correcting errors before data are submitted for use. VA decoder is one of the best error correction mechanisms at the

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

receiving end in a noisy environment [5]. However, VA decoder faces a great challenge when burst errors occur in a received data block [6], [8] and therefore resulting in residual errors or uncorrected errors.

Quality of a decoder depends on industrial requirements. Few transmission errors can be tolerated in favor of the reduction of the code complexity or other factors. When a high quality VA decoder is needed, other mechanisms such as increasing convolutional encoder's constraint length (memory size), applying hybrid code system, code concatenation, code puncturing or code interleaving are used to enhance decoders.

Increase in convolutional encoder's memory size improves Viterbi algorithm error correcting capability. However, when memory size is higher than 10 the algorithm becomes not useful because the decoding process results in excessive delay due to exponential growth of its decoding computational complexity [9].

Ján Poctavek in his study [10] noted that, Hybrid codes that combine convolutional codes with other types of Forward Error Correction or even Automatic Repeat Request algorithms can be used to improve error correction capability. Hybrid codes put the involved codes either in serial or parallel with the convolutional code, T-Y. Chen et al. [11] in their work came with a hybrid code that allowed short convolutional codes to deliver bit error rate performance similar to a long block length turbo code, but with lower latency. However, it is important to note that hybrid codes are applied when individual codes fail to meet the required quality of service.

Code puncturing technique is another alternative which allows an encoder-decoder pair to change their code rates, which alters the code error correction capabilities without changing their basic structure [12], [13]. Code puncturing involves deleting certain code bits. Puncturing convolutional codes was introduced by Cain, Clark and Geist in 1979 [14] and further modified by Hagenauer in 1988 [15]. The application of the codes is flexible in wireless channel. On the other hand, puncturing adjacent bits result in burst error which in turn degrades the binary convolutional encoder/decoder ability to recover lost bits [13].

Another alternative to enable Viterbi decoder combat the effects of multiple and burst errors is the use of an interleaving utility [16], [17], [18], [19]. Interleaving simply involves rearranging data bits from two or more codewords before transmission on the channel [20]. Interleaving process is done by writing data bits row-by-row into a matrix and reading out column-by-column before sending the data over the channel. The reverse process (de-interleaving) is performed at the receiver to get the original arrangement. This process results into each successive bits of any given codeword to have other symbols that belong to other codewords being interleaved. However, it should be noted that interleaving does not decrease the long-term bit error rate but it is successfully in decreasing effects of burst errors in each codeword or data block. Interleaving results in extra delay as the de-interleaving process can only start after all the interleaved data blocks are received. A convolutional interleaver design introduced by Xu Zhuo [21] is reported to have significantly reduced the experienced time delay. However, where the existing decoding algorithms sufficiently meet the required quality of service, then time could be served by doing away with the interleaving functionality in communication systems.

This paper analyses how the received burst errors drives a Viterbi decoder into a disorder state which eventually results in residual errors. It also proposes a technique which enhances the decoder's capability to control its disorder state and thus reduction of residual errors when subjected to burst errors. The rest of this paper is organized in the following manner: Section II of this paper briefly discusses the encoding and decoding process using binary (2, 1, 2) convolutional encoder and the VA decoder. It also analyses how VA decoder behaves when it receives burst error codewords for decoding. Section III discuses the proposed technique which enhance the VA decoder in reducing number of residual errors when it receives burst errors for decoding. Section IV is all about the key observations made and section V is a conclusion to these efforts.

II. Burst Error Decoding Analysis

The convolutional encoding and VA decoding processes will be briefly discussed to give a reader a quick understanding of the processes followed by burst error decoding. Convolutional encoders are finite state machines. Therefore, state and trellis diagrams are used to analyze and describe the concepts.

A. Convolutional codes

Convolutional codes are popular class of coders with memory, where the current output data block from the encoder is not only a function of the current input block but also of other previous data blocks. Binary convolutional codes are commonly defined by three parameters *n, k,* and *m*.

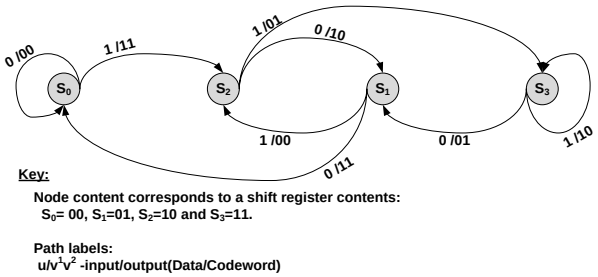


Fig. 2. State diagram of (2, 1, 2) Convolutional encoder

Where *n* is a number of output bits from the encoder and, *k* is a number of input bits to the encoder at a time while *m* is a number of memory size used in that encoder. Therefore, *(n, k, m)* binary convolutional code generates *n* encoded data bits for every *k* data bits, where *(n - k)* is a number of redundancy bits added to data bits.

The encoding process starts from state S_0 and then the encoder branches to different states depending on the data received for encoding. If a *1* bit is received the encoders memory moves from S_0 to S_2. Table I shows the encoding

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

process of *{1-0-1-1-0-0-1-0-0-1-1}* to *{11-10-00-01-01-11-11-10-11-11-01}*.

TABLE I: ENCODING PROCESS

Time Interval	Input bit "u" (data)	Input State	Output bits "v^1v^2" (Codewords)	Next State
1	1	S_0	11	S_2
2	0	S_2	10	S_1
3	1	S_1	00	S_2
4	1	S_2	01	S_3
5	0	S_3	01	S_1
6	0	S_1	11	S_0
7	1	S_0	11	S_2
8	0	S_2	10	S_1
9	0	S_1	11	S_0
10	1	S_0	11	S_2
11	1	S_2	01	S_3

B. Viterbi Algorithm Decoding

VA decoding examines an entire received codeword of a given length at a time interval on a trellis, then computes a metric for each path and makes a decision basing on this metric. One of the frequently used metric for paths evaluation is the Hamming distance metric, which is a bit-wise comparison between the received codeword and the allowable codeword from the decoder. Table II shows the Hamming metric calculation values. The Hamming metrics are computed for each path branch in each time interval and eventually the branch path metrics are cumulatively added to get a total path metric.

TABLE II: HAMMING METRIC CALCULATION

Received C.word	Valid C.word 1	Hamming Metric 1	Valid C.word 2	Hamming metric 2
01	01	2	10	0
00	11	0	00	2
11	01	1	10	1

There are two methods of calculating a Hamming distance metric [9]. In this paper, a method described in Table II where a bit-wise comparison is done and a surviving path is a path with the highest total Hamming metric. Also, at the end the path with highest path metric is considered to be the final path winner. All decoding examples in this paper use this method.

When decoding the received binary convolutional codewords using trellis diagrams, all paths are followed until two or more paths meet on one node. The paths with higher metric are kept and those with lower metric are discarded. When two or more paths have equal high hamming metrics converge on a node then, all the paths are kept and eventually the decoder randomly picks one of those paths. The kept or surviving paths are further repeatedly compared whenever they converge in a node to get a winning surviving path [9], [3]. Fig. 3 shows an example of a trellis diagram of VA in decoding the first four codewords from Table I *(i.e. 11-10-00-01)*, received with a single transmission error in the second codeword. The shaded nodes show the flow of correct decoding path.

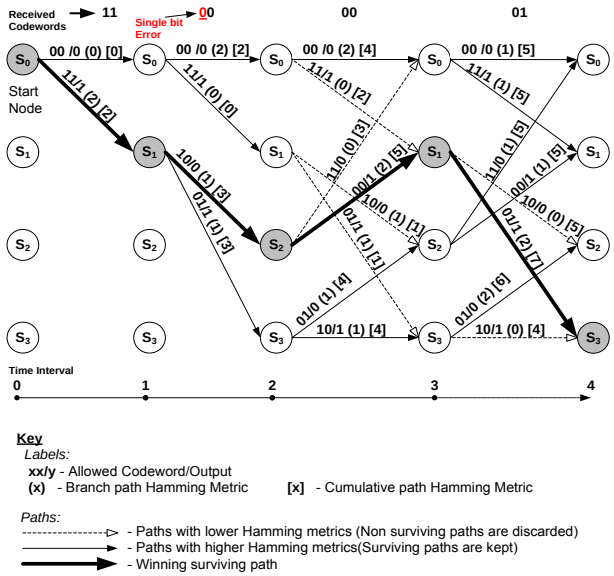


Fig. 3. Trellis diagram of VA decoding

The received codewords (on the top line of Fig. 3) are compared bit by bit with the allowed codewords from the Viterbi decoder using the method described in Table II to obtain the branch metrics. In Fig. 3, path branch metrics are put in path labels using a round bracket i.e. (x). The cumulative path metrics are put in a square bracket i.e. [x]. The decoding process in Fig. 3 follows the following four steps:

- A received codeword is compared with the allowed codeword from the decoder for that particular time interval and the results are put in a round bracket (i.e. (x)) as path branch metric;
- A path's cumulative hamming metric of the current path branch is calculated by adding the obtained branch metric to the cumulative path metric of the immediate predecessor surviving path. It should be noted that, the first path branches from the starting node do not have immediate path branch predecessor, thus, the cumulative branch path metric of their immediate predecessor is zero;
- Compare cumulative path branch metrics of branches which converge in a node. Keep the branch path with higher cumulative branch path metric and consider it as survivor path while other paths are discarded. When the converging paths have the same cumulative path metrics, then they are all kept. Step one to three is repeated at each time interval of the trellis diagram; and
- Finally, the cumulative hamming metrics of all the surviving paths are compared. The surviving path with highest hamming metric is the final winner and data are

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

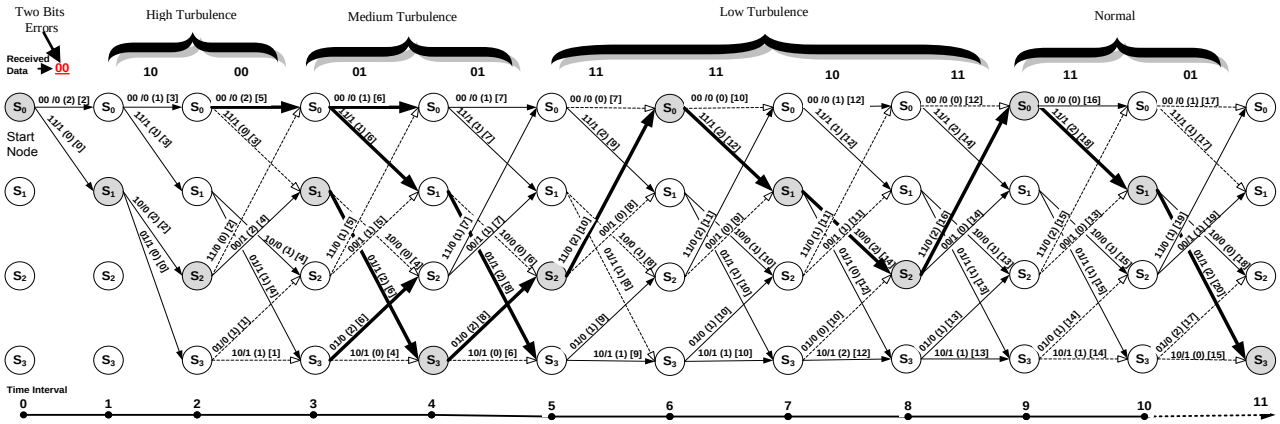


Fig.4 Trellis diagram of Viterbi decoder's turbulence areas

extracted from this path. When more than one path has the same highest cumulative path metrics then one of the surviving paths with highest cumulative path metric is randomly chosen as a final winner.

C. Burst error decoding

In general, binary convolutional decoding algorithms can work well with a single bit transmission error and hardly with spaced multiple bits transmission errors in a data block. VA decoder with a bit decoding rate of 1/2 results in residual errors when it receives more than two bits burst errors in a data block. The reason is that the decoder goes into disorder or turbulence state which persists for about eight consecutive time intervals after receiving only two burst errors. After eight time intervals of no error the decoder resumes to normal state. If the decoder receives even a single error bit in a codeword while in turbulence state then there is a high probability of the decoder to fail and result in a residual error. The turbulence state can further be sub divided into three main categories which are high, medium and low turbulence states depending on how they affect the decoder's decision in the decoding process. Fig. 4 shows division of the turbulence time intervals. The decoder remains in its turbulence state for quite a while before it resumes to the normal decoding state. The decoder's decision unit fails to work correctly when it receives another error while in turbulence resulting in residual errors.

1) High turbulence: *It* occurs only after the decoder receives two consecutive bit errors. This type of turbulence prevails for the next two time intervals of the decoding process immediately after the two burst errors. If the decoder receives another bit error while is in these time intervals then the decoder's decision making unit will fail to follow the correct path because the error lowers the hamming path metric of the correct path and makes it not part of the surviving path and therefore is discarded. Fig. 5 describes the situation in time interval three. The errors obtained in time intervals one and two lowered the hamming distance metric of the correct path branch in time interval three (i.e. from $\mathbf{S_2}$ to $\mathbf{S_1}$) and make it not part of the surviving path.

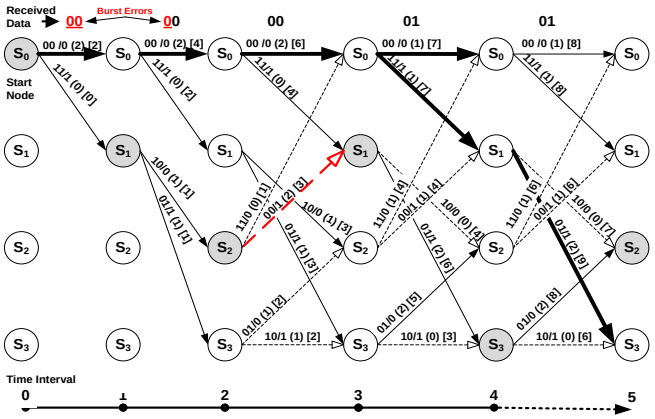


Fig.5. Trellis diagram for high turbulence residual error

2) Medium turbulence: Occurs when the decoder receives another error while it is in medium turbulence area, then the hamming distance of the correct path is lowered and became equal to that of the wrong path. This makes it difficult for the decision unit which decides basing on cumulative Hamming metrics. If this happens, the decision unit randomly picks one path of the two paths creating a possibility of picking a wrong path.

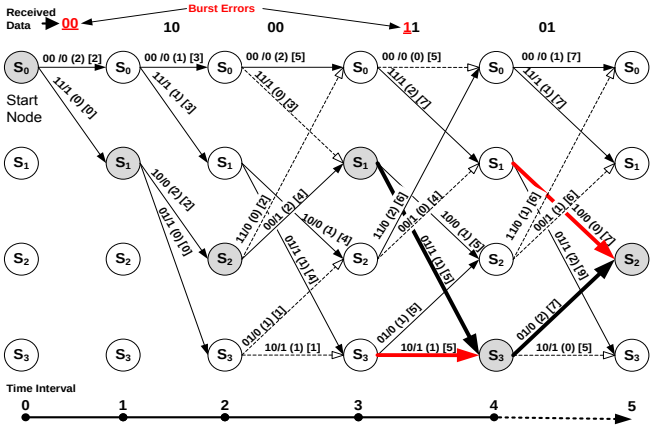


Fig. 6. Trellis diagram for medium turbulence residual error

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Fig. 6 demonstrates how this situation is happening in time interval four and five as result of a single error appeared in time interval four.

3) Low turbulence: Is an error which occurs at the edge or terminal node (end of decoding process) it also lowers the hamming path metric of the correct path and therefore deceiving the decision unit to pick a wrong winning path metric. If this happens persistently residual errors will be experienced. This is because there is little possibility that the surviving paths will meet in the middle except at the starting node. It is important to note that, this type of error can also occur in the high and medium turbulence areas if they happen to be at the terminal node. Fig.7 shows an example of how this error can occur at the edge of decoding. Since decoders are sometimes forced to terminate at S_0 state, then tail bits can reduce the possibility of this error to occur.

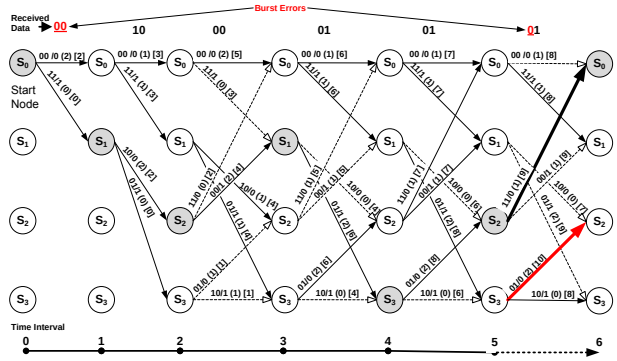


Fig. 7 Low turbulence residual errors

III. Enhanced Viterbi Algorithm

After describing how burst errors cause havoc to Viterbi decoder, this part of the paper introduces to readers an enhancement to the encoding and decoding procedures that improves the decoder's operational performance when a 1/2 VA decoder receives two bits burst errors and an error occurs in the medium and low turbulence areas.

The proposed solution includes locking the encoder by adding either two low bits *(i.e. 00)* or two high bits *(i.e. 11)* after each data bit to be encoded at the sender's machine

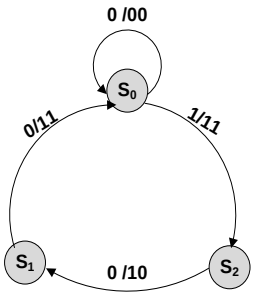


Fig. 8 State diagram of a lower locked (2, 1, 2) Convolutional encoder

The two lock bits forces the encoder to work either on the lower end side of the encoder or the higher end side of the encoder but not both (see Fig. 2). Assuming the following data stream of bits *{1, 0, 1….}* is to be encoded. Locking the encoder by using two lower bits is just adding two zero bits after each bit to be encoded. Therefore, the bit stream to be encoded will look like *{1-0-0-0-0-0-1-0-0…}*. If we want to lock the encoder using two high bits the bit stream will look like *{1-1-1-0-1-1-1-1-1…}*. Fig. 8 shows the lower end locked (2, 1, 2) binary convolutional encoder and how it works. Table III shows the enhanced encoding process of the mentioned string of bits *{i.e. 1, 0, 1…}* using a lower locked encoder and their corresponding output codewords in relation to the encoder's state transitions.

Lock bits are also transmitted with data in the transmission channel. This fact lowers the encoder's data transmission rate from 1/2 to 1/6 which means for each data bit there are six bits to be transmitted in channel. This is considered to be one of the tradeoff of this method. Lower locked encoder ignores state S_3 completely and works perfectly with the remaining three states. This feature gives the VA decoder special characteristic which enable it to use Non Transmittable Codewords (NTCs) at the receiving machine. Table III shows the enhanced encoding process where in time interval one, four and seven are data bits and the rest are lock bits for lower end locked convolutional encoder.

TABLE III: ENHANCED ENCODING PROCESS

Time Interval	Input bit "u" (data)	Input State	Output bits "v^1v^2" (Codewords)	Next State
1	1	S_0	11	S_2
2	0	S_2	10	S_1
3	0	S_1	11	S_0
4	0	S_0	00	S_0
5	0	S_0	00	S_0
6	0	S_0	00	S_0
7	1	S_0	11	S_2
8	0	S_2	10	S_1
9	0	S_1	11	S_0

NTC is either two zero-zero bits (i.e. 00) for lower end locked encoder or two one-one bits (i.e.11) for higher end locked encoder. NTCs can be added to the received codewords before they are submitted to VA decoder for decoding. Unlike lock bits, NTCs do not decrease the encoder's data transmission rate as they are not transmitted over the channel. Since NTCs are known codewords *{i.e. 00-00-00… for lower locked encoder}*, then they can be easily added to the received data codewords before they are submitted for the decoding process. After the decoding process, bits corresponding to the added NTCs and lock bits are discarded and the remaining data are submitted for use. Suppose adequate NTCs (let say six codewords) to cover the medium and low turbulence areas of fig.4 were added as prefixes to each received data codeword, then results shown in fig. 9 could be obtained,

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

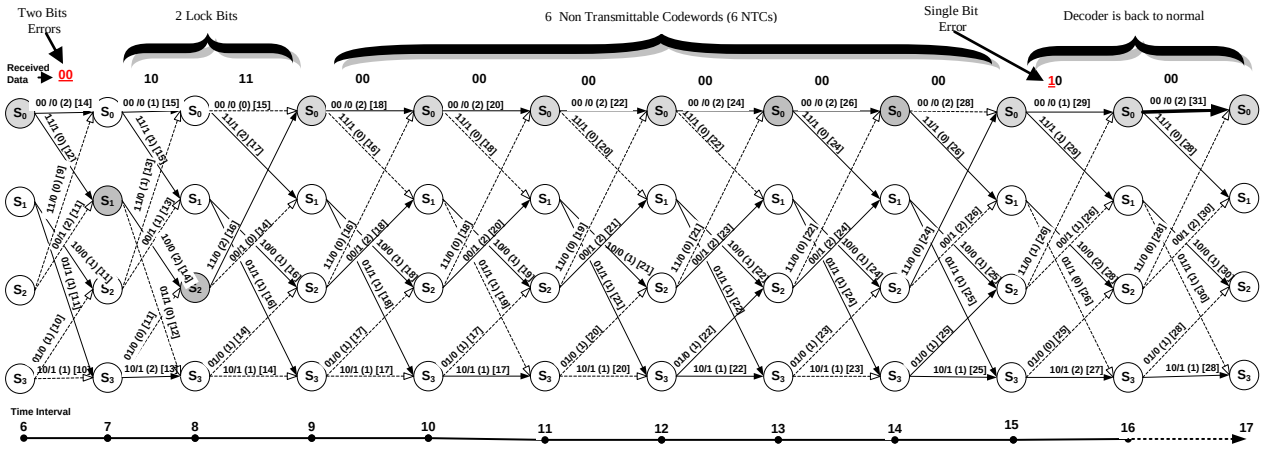


Fig. 9 Truncated trellis diagram for 6 NTCs Enhanced Viterbi Decoder

where time interval one to six (truncated) and ten to fifteen are the NTCs added as prefixes to each received data codeword, time interval seven and sixteen are data codewords and time interval eight and nine are lower locking codewords received. The six NTCs are enough to push the transmitted data codewords out of the medium and low turbulence areas. Fig. 9 shows how this fact reduces to zero the error occurrence possibility in these areas.

After the encoding process as per table III, both codewords corresponding to data and lock bits are transmitted through channel. All transmitted codewords are susceptible to noise and distortion in the channel. Let us assume that the transmitted codewords corresponding to lock bits are safe and the first two bits corresponding to the first data bit are distorted as well as the first bit of codeword corresponding to the second data bit. Hence, there will be two burst errors and another single error that could fall on the medium turbulence area of the decoding process. Since the six codewords were added to the received sequence before decoding, then, the single error bit is pushed out of the medium and low turbulence areas where it could cause a residual error or errors. After the addition of the six NTCs the decoder resumes to its normal state. When the decoder resumes to normal state it can easily decode a single or multiple errors again. Fig. 9 is a truncated trellis diagram that shows the location of two lock bits and six NTCs in relation to data bits when stabilizing the decoder to normal. The enhanced decoder successfully solves a residual error problem which could occur in time interval ten (medium turbulence area) by delaying it until the decoder is back to normal in time interval seventeen of the decoding process. Note that time intervals one to six are truncated from the fig. 9 for the easy demonstration.

NTCs can be added as one codeword, two codewords and so on. However, if we consider a 1/2 Viterbi decoder an increase of the number of NTCs from one to six is expected to have an increasing impact at decreasing rate of reduction of the number of residual errors. Any addition of NTCs after six NTCs is expected to have less impact on decreasing the number of residual errors because the decoder has already resumed to normal.

Medium and low turbulence of 1/2 viterbi decoder prevails for six time intervals out of all eight time intervals of turbulences. Therefore, it can simply be logically deduced that a VA decoder decoding binary data from a locked convolutional encoder can reduce decoding residual errors by 75 percent if supplied with sufficient NTCs that covers its medium and low turbulence areas.

IV. Remarks and Discussion

Through the description of the occurrence of residual errors due to the reception of burst errors during VA decoding and the proposed remedy, a number of factors affecting operational performance of Viterbi decoder in burst errors and the proposed solution have been observed and noted as follows:

- Burst errors drive VA decoder into a confusion state which was called turbulence. The turbulence area have been further subdivided into high, medium and low turbulences basing on the way they cause residual errors;
- A VA decoder in high turbulences causes a residual error if it receives another erroneous codeword;
- A viterbi decoder in medium or low turbulences may also cause residual error or errors if it receives another erroneous codeword;
- Locking a convolutional encoder using binary bits enabled it to use the proposed Non Transmittable Codewords (NTCs) at the receiving machine;
- Locking bits lowers the convolutional encoder's transmission rate as they are also transmitted;
- NTCs are known Codewords to the data receiving machines to enable the system work perfectly;
- NTCs are added to the received data codeword before they are submitted for decoding process. Therefore, NTCs are expected to be error free as they are not transmitted and do not lower the convolutional encoder's data transmission rate;

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

- Bits corresponding to NTCs and lock bits codewords are removed after the decoding process, the remaining data are submitted for use; and
- Sufficient NTCs that cover the VA decoder's medium and low turbulence areas completely solve the problem of residual errors that may occur in those areas.

V. Conclusions & Recommendations

This paper analyzed and discussed the issue of burst errors that lead to unbearable amount of residual errors in VA decoding. It also proposes the locking convolutional encoder technique that allows the use of Non Transmittable Codewords (NTCs) to be applied in enabling VA decoders to reduce the amount of residual errors in bad or poor channels. The proposed technique shows the possibility of reducing by 75 percent of residual errors occurring in VA decoders when sufficient NTCs are used with a locked encoder. However, for the (2, 1, 2) binary convolutional encoder the technique lowered the encoder's data transmission rate from 1/2 to 1/6 as the encoding lock bits need to be transmitted as well. Simulation of the proposed technique using appropriate simulation software to show the improvement that the enhanced VA decoder introduces in recovering of transmitted data under different signal to noise ratio is being recommended.

Acknowledgment

This work is part of the first author's PhD work, which is supported by Tanzania Commission for Science and Technology (COSTECH), The Nelson Mandela African Institution of Science and Technology (NM-AIST) and The University of Dodoma (UDOM).

References

[1] A. Viterbi, "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm," *Information Theory, IEEE Transactions on*, vol. 13, pp. 260-269, 1967.

[2] S. I. Mrutu, S. Kalolo, M. Byanyuma, C. Nyakyi, and A. Sam, "Bandwidth Aware Fec Algorithms for Wireless Communication Systems," *Control Theory and Informatics*, vol. 3, pp. 8-13, 2013.

[3] A. Viterbi, "Convolutional Codes and Their Performance in Communication Systems," *Communication Technology, IEEE Transactions on*, vol. 19, pp. 751-772, 1971.

[4] J. Omura, "On the Viterbi Decoding Algorithm," *Information Theory, IEEE Transactions on*, vol. 15, pp. 177-179, 1969.

[5] D. Forney Jr, "Convolutional Codes Ii. Maximum-Likelihood Decoding," *Information and control*, vol. 25, pp. 222-266, 1974.

[6] S. Nouh, I. Chana, and M. Belkasmi, "Decoding of Block Codes by Using Genetic Algorithms and Permutations Set," *International Journal of Communication Networks and Information Security (IJCNIS)*, vol. 5, pp. 201-209, 2013.

[7] B. Forouzan, C. Coombs, and S. C. Fegan, *Introduction to Data Communications and Networking*: McGraw-Hill, Inc., 1997.

[8] W. FS FILHO and D. S. E. S. EH, "Ea,"Adaptive Forward Error Correction for Interactive Streaming over the Internet"," in *IEEE Globecom'06*, 2006, pp. 1-6.

[9] S. I. Mrutu, A. Sam, and N. H. Mvungi, "Forward Error Correction Convolutional Codes for RTAs' Networks: An Overview," *International Journal of Computer Network and Information Security*, vol. 6, pp. 19-27, 2014.

[10] J. Poctavek, K. Kotuliaková, J. Polec, M. Osadský, and S. Ondrušová, "Throughput Parameter Optimization of Adaptive Arq/Harq Scheme," *International Journal of Communication Networks and Information Security (IJCNIS)*, vol. 3, pp. 89-95, 2011.

[11] T.-Y. Chen, N. Seshadri, and B.-Z. Shen, "Is Feedback a Performance Equalizer of Classic and Modern Codes?," in *Information Theory and Applications Workshop (ITA), 2010*, 2010, pp. 1-5.

[12] H. Liu, H. Ma, M. El Zarki, and S. Gupta, "Error Control Schemes for Networks: An Overview," *Mobile Networks and Applications*, vol. 2, pp. 167-182, 1997.

[13] L. Sari, "Effects of Puncturing Patterns on Punctured Convolutional Codes," *Telkomnika*, vol. 10, pp. 759-770, 2012.

[14] J. Cain, G. Clark, and J. M. Geist, "Punctured Convolutional Codes of Rate (N-1)/N and Simplified Maximum Likelihood Decoding (Corresp.)," *Information Theory, IEEE Transactions on*, vol. 25, pp. 97-100, 1979.

[15] J. Hagenauer, "Rate-Compatible Punctured Convolutional Codes (Rcpc Codes) and Their Applications," *Communications, IEEE Transactions on*, vol. 36, pp. 389-400, 1988.

[16] I. Jacobs, "Practical Applications of Coding," *Information Theory, IEEE Transactions on*, vol. 20, pp. 305-310, 1974.

[17] H. Zerrouki and M. Feham, "High Throughput of Wimax Mimo Ofdm Including Adaptive Modulation and Coding," *International Journal of Computer Science and Information Security(IJCSIS)*, pp. 86-91, 2010.

[18] F. Escribano and A. Tarable, "Interleaver Design for Parallel Concatenated Chaos-Based Coded Modulations," *IEEE COMMUNICATIONS LETTERS*, vol. 17, pp. 834-837, 2013.

[19] G. Wang, A. Vosoughi, H. Shen, J. R. Cavallaro, and Y. Guo, "Parallel Interleaver Architecture with New Scheduling Scheme for High Throughput Configurable Turbo Decoder," in *Circuits and Systems (ISCAS), 2013 IEEE International Symposium on*, 2013, pp. 1340-1343.

[20] T.-Y. Chen, N. Seshadri, and R. D. Wesel, "Incremental Redundancy: A Comparison of a Sphere-Packing Analysis and Convolutional Codes," in *Information Theory and Applications Workshop (ITA), 2011*, 2011, pp. 1-5.

[21] Z. Xu, S. Guan, and F. Yao, "A Novel Low-Time-Delay Convolutional Interleaver and Its Performance," in *Information and Communications Technologies (IETICT 2013), IET International Conference on*, 2013, pp. 208-212.

Authors Profile

Salehe I. Mrutu received his B.Sc. and M.Sc. degrees in Computer Science from The International University of Africa in 2003 and the University of Gezira in 2006 respectively. He is currently a Ph.D. student at the school of Computational and Communication Science and Engineering (CoCSE) of the Nelson Mandela African Institution of Science and Technology (NM-AIST) in Arusha, Tanzania. He is also serving as assistant lecturer at the University of Dodoma (UDOM) under the school of informatics since 2007. His research interests include Forward Error Correction (FEC) codes, quality-of-service provisioning and resource management for multimedia communications networks.

Anael Sam received his B.Sc., M.Sc. and Ph.D. in Electronics Engineering (Institute of Electronics and Photonics, Slovak University of Technology, Slovak Republic). He works as senior lecturer at the Nelson Mandela Institution of Science and Technology, Arusha, Tanzania. Dr Sam's specialization and research interests are in radio, multimedia and mobile communication systems; electronics and telecommunication engineering, software quality assurance engineering and mobile networks optimization. He is also a member of IEEE and ISQTB international.

Nerey H. Mvungi received the B.Sc. degree in electrical engineering from the University of Dar Es Salaam, Tanzania, in 1978; the M.Sc. degree in electronics control from Salford University, U.K. in 1980; and the Ph.D. degree from Leeds University Leeds, U.K. in 1989. He worked for a year with the Phillips Center for Technology, Eindhoven, Eindhoven, and the Netherlands February 1992 to Feb 1993. He was attached to Onersol Solar Energy Research Centre in Niamey June-July 1991 as ILO Consultant on Solar Energy Systems.

Since his undergraduate graduation in 1978, he has worked as an academician and is now a full professor. He has mostly worked in the University of Dar es Salaam but for the period of September 2008 to June 2012 when he was at the University Dodoma in Tanzania to starting a new IT College as its founding Principal.

Prof. Mvungi's research interests are in control and instrumentation, computer communication and applied electronics, lightning protection, rural access, power-quality aspects, and remote monitoring and control of energy consumption and digital broadcasting. He received a 2010 IBM Faculty Award.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Towards a Mobile-Based DSS for Smallholder Livestock Keepers: Tanzania as a Case Study.

Bernard Mussa*, Zaipuna Yonah,
Computational and Communication Science and Engineering,
The Nelson Mandela African Institution of Science and Technology,
Arusha, Tanzania.

Charles Tarimo,
College of Engineering and Technology,
University of Dar Es Salaam
Dar Es Salaam, Tanzania.

Abstract--**Building a useful and responsive Decision Support System (DSS) requires a deep understanding of the pertinent application domain before starting the system design. In this paper we report about an attempt to develop a mobile-based DSS for smallholder livestock keepers with Arusha region as a case study. The objective of the reported study is to provide an information tool for decision making to the smallholder livestock keepers. The development process involved: 1) employing information gathering techniques to understand smallholder livestock keepers' information needs 2) studying the current methods that are used for information flow among livestock stakeholders. (i.e. smallholder livestock keepers, extension officers and livestock researchers) 3) analysis of the current situation within Arusha: located in the northern parts of Tanzania in terms of mobile phones penetration, with prospects of leveraging the high mobile phone penetration rate for enhanced information sharing among the smallholder livestock keepers and 4) exploration of options for the platform/model to be used for information access and delivery. The outputs of the above four activities were used to inform the requirements elicitation, and design phases of the mobile-based DSS system development. In addition, the mentioned four activities were supplemented by an extensive literature review of related works on requirements engineering in DSS development. It is anticipated that once the system has been developed, it will be of help to livestock keepers, improving farm-level productivity and decision making process. Findings from the study indicate that majority of smallholder livestock keepers in the selected area possess mobile phones and are in need of access to specific information to support their livestock related decision making. However, information access platforms/models that are currently in place do not cater for a satisfactory solution to their needs. Analysis of various options for designing a DSS platform has shown that a model that considers the administrative, organizational structure, as well as roles of relevant stakeholders in the livestock information flow will be useful for the studied context. The proposed Role-based Information Decision Support (RIDS) Model will facilitate data querying, analysis and information delivery based on users' information requirements for the design of the DSS's data marts. This will, in turn, be the basis for implementing a system of information sharing and delivery mechanism that will improve the decision making process and livestock management for smallholder livestock keepers in the studied geographical environment.**

Keywords: Decision Support System, Data Mart, Mobile phones, Smallholder livestock keepers.

I. INTRODUCTION

A Decision Support System (DSS) for information retrieval, data analysis and decision making support can be a useful tool for enhancing the productivity of livestock sector as reported in [1]. In our case study, despite the fact that data exists from different livestock data sources, smallholder livestock keepers rarely access this data for their decision making. Obviously, information is required to support decisions making by individuals and organizations if they are to remain more competitive and productive.

Data is a valuable asset and represents a tremendous investment of resources. There are unprecedented volumes of data today existing in a variety of places and different formats. The growing volume of data has sparked renewed interest in data analysis [2], thus making it imperative to have some techniques for data integration and analysis so as to provide a linkage between data collection and potential use.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

In the quest of developing a software system such as a DSS, goals and users requirement must be identified as an initial step towards building a complete system. The idea of user-centric approach is very pertinent towards development of any information system. As for developing an effective DSS using data warehousing techniques, it is pointed out by Rai *et al.* in [3,20], that, user requirements play a fundamental role in restricting the area of interest for data analysis and in choosing facts, dimensions, and measures for data marts that are to be designed and implemented.

The objective of the study reported in this is to develop an effective mobile-based DSS that is responsive to its intended users—smallholder livestock keepers with those in Arusha as a case study. Apparently this endeavor calls for the identification of proper information needs of the said smallholder livestock keepers and availability of an adequate platform/model for access and delivery of such relevant and in-demand information. Using our case study area, which is the Meru District in Northern Tanzania, we have identified four (4) key user groups, namely: district livestock officers, ward livestock field officers, smallholder livestock keepers and livestock researchers that play important roles in the livestock data and information exchange within the district, and on which the requirements elicitation process has been focused.

Weibelzahl *et al.* in [4] remarked that, involving users from the very beginning can help to discover their mental models and expectations, to identify and analyze theoretical tasks, workflow and goals, and in general to validate the developers' assumptions about the users.

Users' and system requirements for DSS design that are presented in this paper are results of analysis of data obtained from interviews, questionnaires, document reviews and group discussions with key user groups identified in this study. These will serve as a guide to business, functional and non-functional information requirements for the developer of the mobile-based DSS for livestock keepers in Tanzania.

From the analysis of data collected, majority of smallholder livestock keepers depend on information delivered by livestock experts around and within their administrative locality. Based on this fact, we propose a Role-based Information Decision Support (RIDS) Model as suitable for meeting information needs of smallholder livestock keepers. The proposed model will dictate the design of specified data marts for enhancing data analysis and information delivery based on users' information needs and mobile capabilities. The model considers the administrative, organizational structure as well as roles of relevant stakeholders in the livestock information flow in the Tanzanian context and is designed according to flow of information from the source of information (source systems), granular analysis of data and information dissemination agents/middleware to end-user of information for decision making support.

This paper is organized in 5 sections. Section one covers general introduction of the research topic and objective of the study. Literature review and related works are discussed in section two. Section three covers Methodology employed in this research work. Results and Discussion of analysis of the data collected are covered in section four and the paper ends with a conclusion in section five.

II. LITERATURE REVIEW

Recent experiences in building Decision Support Systems (DSS) point out the need of a deep understanding of the application domain before starting a system design. The application domain under consideration has to be characterized in terms of stakeholders' roles and of their requirements and in terms of the decision making processes these stakeholders are involved in [5].

In designing a DSS using Data Warehousing (DW) techniques it is necessary to distinguish between supply- and demand-driven approaches. Inmon [6] describes a supply driven approach in development of DSS as opposed to requirement–driven development

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

of operational systems. In demand driven, users' information needs are given more relevance.

A goal oriented approach to requirement analysis is proposed by Giorgini *et al*. [7], in which two perspectives are integrated for requirement analysis: organizational modeling, centered on stakeholders, and decisional modeling, focused on decision makers. The approach used is similar to proposed one but differs in the design in which the former relies on the organizational modelling while ours is centered on end-users who are the targeted decision makers.

DSS requirements are identified in terms of goal and plan delegation from stakeholders (the users) to the system-to-be in an agent-oriented software engineering methodology proposed by Perini et al. in [5]. In this, early requirement analysis process is analogous to the one used this paper but there is a significant difference when it comes to mapping of available supply information to users informational requirements.

Anton [8] suggested a goal-based requirement analysis whereby the goal analysis to identify requirements is more on the organizational goals as compared to users' goals. Here, the approach described favors the information needs of organizations rather that the end-users of the DSS which is the main focus employed in the proposed user-driven approach.

An interesting case-based comparison of supply- and demand-driven approaches that is worthy to mention can be found in [9, 14]. Extraordinarily, it is concluded that data-oriented and goal-oriented techniques are complementary, and may be used in parallel to achieve optimal design.

Finally, it is worth to mention that related works in requirements analysis have all stressed on the user involvement in the early stages of DSS development. Studies have also shown that 40% of all DW projects are never completed, and 85% fail to meet business objectives [10] reasons behind being failure to accurately collect and analyses requirements.

The new proposed approach to requirement elicitation is mainly user-driven. Also adopted is the mixed demand/supply mechanism in the requirement analysis whereby information needs of users are mapped and fulfilled as per supply of available data in the operational system databases. This approach is both cost effective as well as saves system development time.

III. METHODOLOGY

The research methodology employed in the reported study was based on qualitative research methods such as interviews, observations, questionnaires, documents analysis, participating in group discussions related to the research topics, literature review and analysis of existing systems.

By the term qualitative research, Strauss and Corbin [11] defines it as, "any kind of research that produces findings not arrived at by means of statistical procedures or other means of quantification".

A. Demographics of Respondents

A total of 210 smallholder livestock keepers were purposely and conveniently sampled to represent the population in Meru District in Arusha Region-Tanzania. The location was selected due to concentration of smallholder livestock keepers in the region. The study sample comprised of 108 (51.43%) female and 102 (48.57%) males and respondents were distributed across the administrative wards in the district as shown in Fig. 1.

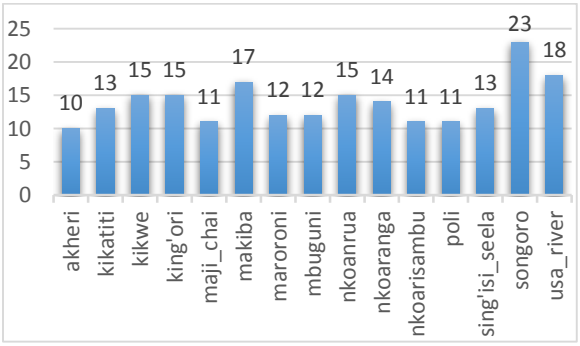


Figure. 1. Respondents Distribution by Ward

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

In addition, Thirty (30) livestock experts comprising of Fifteen (15) Ward Livestock Field Officers from respective wards, Ten (10) District Livestock Officers from Meru District Council Office and Five (5) Livestock Researchers from livestock training and research institutions namely the Nelson Mandela African Institution of Science and Technology (NM-AIST) and Livestock Training Agency (LITA-Tengeru) were also key stakeholders involved in the reported study.

B. Case Study Area Mapping

The use of ODK tool for data collection enabled the gathering of GPS location information of respondents during the process of requirement collection. Geographical locations of the interviewed livestock keepers were, with their consent, recorded and plotted on Google App Engine Maps Visualizer as shown in Fig. 2. The dispersed location distribution of livestock keepers highlights the importance of exploiting mobile technology for information dissemination to targeted users since one of the challenges pointed by the wards' livestock field officers in relation to information dissemination is the geographical remoteness of livestock keepers.

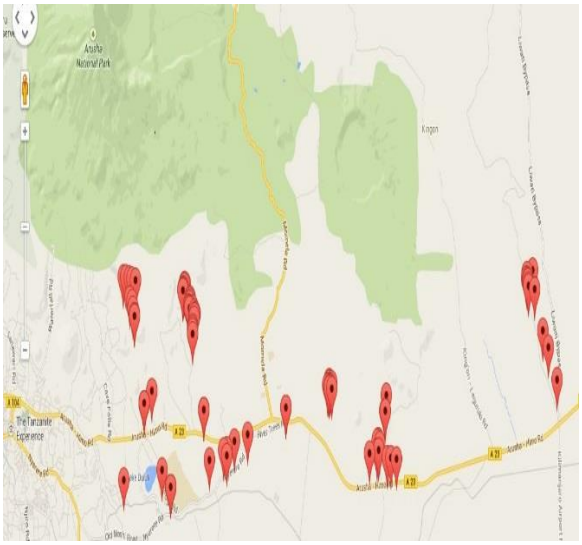


Figure. 2. A map showing the study site

C. Requirements Elicitation Methods

a. Information gathering process (identification of information needs)

Questionnaires, Group discussions and Interviews were used in gathering requirements with regards to information and decision making support needs for the development of a Decision Support System (DSS) for the targeted user groups namely smallholder livestock keepers**,** district livestock officers, ward livestock field officers and livestock researchers.

The questionnaires were specifically set to investigate information needs of the respondents, current situation with regards to information flow among livestock stakeholders as well as mobile phones penetration rate in the studied area. Guided and self-response questionnaires were administered to the targeted groups. For smallholder livestock keepers, structured questionnaires were designed and administered using a guided interview through an Open Data Kit (ODK) tool. ODK is free and open source suite of tools that allow data collection using mobile devices and data submission to an online server, even without an Internet connection or mobile carrier service at the time of data collection [12].

Facilitated group discussions involving wards livestock field officers and livestock keepers were also conducted for the purpose of understanding stakeholders' roles, interactions and information flow among them.

Interviews were conducted with livestock researchers and some livestock officers in order understand from the experts' point of view on how research findings and relevant information deemed essential could be exploited to directly serve information needs of smallholder livestock keepers through the proposed DSS.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

b. Analysis of existing data flow and source systems

Detailed review and analysis of relevant documents obtained from wards' and district council offices was carried out in order to identify data, processes and tools that are currently being employed at various levels and understand the organizational structure with respect to data collection, information exchange as well as identifying existing operational source systems that could serve as data sources for the DSS to be developed.

A demand/supply driven approach was adopted whereby users' information demands obtained through analysis of data collected were analyzed in parallel with identified existing operational systems (i.e. Livestock Database System) in order to find the availability of data that can be used to address the information requirements of smallholder livestock keepers. This method was used because information demands of the users could mainly be fulfilled by data that were provided by the existing operational source systems at the district office. Fig. 3 below shows the processes involved in the Demand /Supply model adopted.

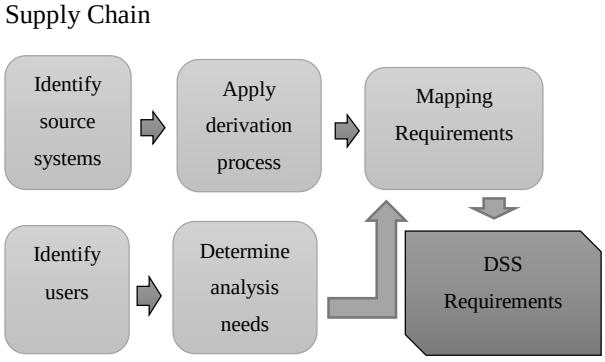


Figure. 3. Demand/ supply driven requirement gathering and analysis

TABLE I: TASKS AND TECHNIQUES USED IN THE SUPPLY/DEMAND DRIVEN APPROACH

Resources / **Task**	**Technique**
Identify source systems	Interview, Observation
Apply derivation process	Reverse Engineering of existing schema(Livestock Database System)
Identify users	Interview, Documents Review, Group Discussions
Determine analysis needs	Documents Review, Literature review
Mapping requirements	Design and Literature Review
DSS requirements specification	Use cases, Documentation

IV. RESULTS AND DISCUSSIONS

A. Smallholder livestock keepers' information needs

a. Information requirements from smallholder livestock keepers

In order to capture information needs of livestock keepers, who are the primary target user group for the DSS to be developed, a total of 210 questionnaires were administered. These were focused on the kind of information needed to support decision making. Table 2 summarizes respondents' information needs. These results showed a significant number of livestock keepers are in need of information on disease outbreaks, vaccinations and treatment, markets and weather information and modern methods of livestock husbandry. This variety of targeted information will support their daily decision making regarding their livestock and the overall livestock keeping process.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

TABLE II: PREFERRED INFORMATION NEEDS OF LIVESTOCK KEEPERS FOR DECISION MAKING SUPPORT

	Information Type	Number of Responses	Percentage (%)
Smallholder Livestock Keepers Information Needs	Market Information	190	90.5
	Disease Outbreaks	200	95.2
	Vaccinations & Treatment	180	85.7
	Weather	150	71.4
	Modern Livestock Husbandry	200	95.2

b. Information requirements as per livestock experts

For the purpose of effective users' requirements elicitation, livestock experts were also asked on what information they deem necessary for livestock keepers to improve their productivity. The experts included district and ward level extension officers as well as livestock researchers who have field experience on their working domains. These experts play an important role to ensure livestock keepers are fed with not only accurate but also reliable information for improving their productivity.

A total 30 questionnaire were administered to livestock experts in this aspect. It is noted that, 96% of the respondents suggested that livestock market information, weather information and disease related information present an essential opportunity for livestock keepers to improve their productivity as summarized in Table 3. Findings indicate that, such information is usually sought by smallholder livestock keepers when in consultation with these experts.

TABLE III: INFORMATION RECOMMENDATION BY LIVESTOCK EXPERTS

	Information Type	Number of Responses	Percentage (%)
Livestock Experts Information Recommendations	Market Information	29	96
	Disease Information	29	96
	Weather Information	29	96

In view of the foregoing, the implication is that data available from the existing data sources located at the district livestock office can be properly analyzed and the outcomes subsequently disseminated in order to meet the information needs of target users.

B. The current methods that are used for information flow among livestock stakeholders

Results from analysis of collected data indicate that information among livestock stakeholders is currently being shared through informal and social interaction that usually depend upon the chance that such social meetings will occur. With 82% of smallholder livestock keepers relying on information from livestock field officers in their locality mainly through face to face meeting, the efficacy of this method is currently being hindered by geographical remoteness of recipients' homesteads especially in rural areas.

Furthermore, findings show that mass media communication such as radio, TV provides another channel for information flow in the current situation. Livestock researchers have to coordinate with district offices to obtain or provide information to livestock keepers.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

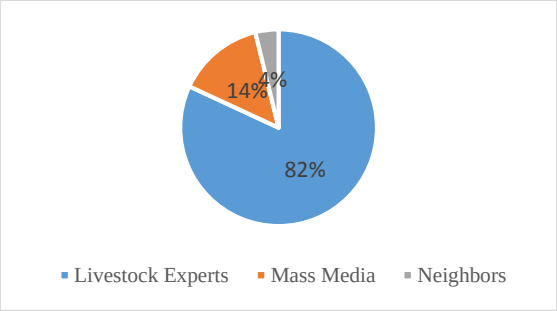


Figure. 4. Smallholder livestock keepers' sources of information

Additionally, livestock data are collected monthly by livestock field officers in respective wards and analysis of such data is only done at the district level and no relevant feedback sent to livestock keepers.

C. *Mobile phone penetration and usage*

Findings show that 98% of the respondents possess and are conversant with mobile phones usage implying that access to relevant farm information may be improved through mobile phone technologies. Furthermore, 97% of these prefer mobile phones as a media for information delivery and communication with livestock officers. This emphasizes the need for the implementation of a mobile based DSS for livestock keepers.

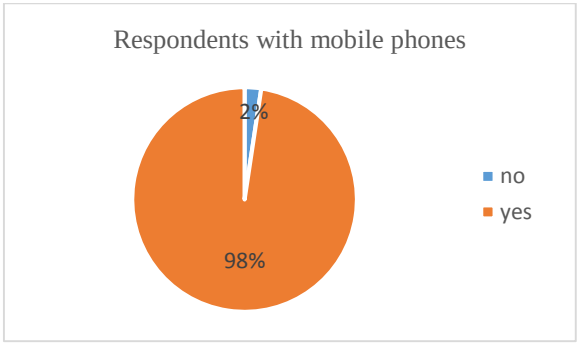


Figure.5. Mobile Phones Penetration

Potential DSS Users and their Roles

Potential users of the DSS were identified and categorized as in Table 4.

TABLE IV: POTENTIAL DSS USERS

User group	**District Livestock Officers (DLO)**	**Ward Livestock Field Officers (WLO)**	**Smallholder Livestock Keepers (SLK)**	**Livestock Researchers (LR)**
Roles	Source of DSS data and information. (Operational Source System)	Use the DSS to disseminate information to livestock keepers	Targeted users. Use the DSS to query and receive information	Use the DSS to analyze data for researchable problems and disseminate relevant research findings via wards and district offices

D. *Platform/model to be used for information access and delivery*

Due to the nature of data collection, storage and information retrieval that exist in the information cycle among the concerned stakeholders identified in this study, it is imperative to consider each role in the cycle and how these roles can be leveraged to ensure proper decision making support to livestock keepers' through provision of relevant and in-demand information. Thus, in contextualizing the development of the DSS to get the best decision making support, we propose use of Role-based Information Decision Support (RIDS) Model that will facilitate data querying, analysis and information delivery based on users' information requirements and roles. The model is depicted in Fig. 6:

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

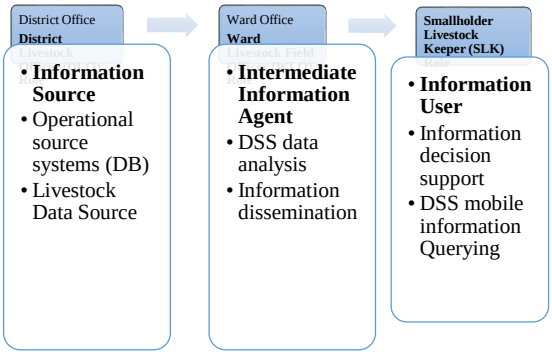


Figure. 6:.Role-based Information Decision Support (RIDS) Model

The description of the model is as follows:

a) Information source (district livestock officer role)

The district office is the main source of livestock data and information. Data collected by livestock field officers from villages is aggregated at the ward level and then sent to the district office on a monthly basis. At the district level, data from all the wards in the respective district is entered in the Livestock Database System for analysis and use at the district level. Development of a DSS for livestock keepers should consider the role played by the District Livestock Officer with respect to ensuring livestock data is made available for appropriate data analysis so that information can be sent to respective ward's livestock field officer. This will enhance availability of source data and facilitate extraction of such data to be utilized in the DSS.

b) Intermediate information agent (ward livestock field officer role)

At this level, the information received is semi-analyzed in the sense that data is now at the ward level of granularity. The Intermediate Information Agent (IIA), the role of wards' livestock field officer, is to ensure relevant information is disseminated to livestock keepers in the respective ward. This is the intermediate recipient of data from the district level as well as the main source of information to livestock keepers. In designing the DSS, the IIA is an important role due to the fact that livestock keepers rely much on the information provided by these experts and are in direct contact with them for the decision making regarding their livestock. This should be considered as the main role in delivering information to target users of the DSS and much of the requirement of such a system should build upon the existence of the information agents as the first line of communication and information exchange between livestock keepers and other concerned stakeholders.

c) Information user (smallholder livestock keeper role)

The model also recognizes the role of Information Users, who are the livestock keepers themselves. The Delivery of information to these target users has taken into account the current exchange of information among users and providers. Findings show that mobile phones have been used in delivering information to livestock keepers from livestock field officers implying that the intermediate level of the model could facilitate information delivery to target users after appropriate querying of data from the source systems.

This model is the proposed guide towards successful implementation of the mobile-based DSS for livestock keepers in Tanzania. It reiterates the need to focus on the source of data, information requirements, data analysis requirement, information sharing as well as information organizational structure that considers the underlying information flow in the application domain.

V. CONCLUSION

The availability and use of mobile phones among smallholder livestock keepers in Tanzania present the most favorable possibility of implementing a mobile-based Decision Support System for the smallholder livestock keepers. In this paper, we have presented information requirements of the users obtained through user-driven requirements elicitation approach. We also propose a Role-based Information Decision

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Support model for facilitating data querying, analysis and information delivery, which has also considered administrative organizational structure of the current information flow among the concerned stakeholders. The advantage of the presented approach is that it ensures that early user requirements are properly taken into account; hence ensuring a good design that is responsive to the anticipated user needs. The proposed model for information access and delivery presented here forms a deliverable from the first phase our journey towards the envisaged Mobile-Based DSS for Smallholder Livestock Keepers in Tanzania; whereby this phase mainly covers the early requirement gathering Process and the Actors. It is upon this (the proposed) information access and delivery model that the said mobile based- DSS is going to be built in the next phase—the system design and implementation phase.

ACKNOWLEDGEMENTS

The authors express their appreciation to Nelson Mandela African Institution of Science and Technology (NM-AIST) for providing financial and material support for the research undertaking. Support provided by the Meru District Council Livestock Office is also recognized and acknowledged.

REFERENCES

[1] Castelan-Ortega, O. A., Fawcett, R. H., Arriaga-Jordán, C., & Herrero, M. (2003). A Decision Support System for smallholder campesino maize–cattle production systems of the Toluca Valley in Central Mexico. Part I— integrating biological and socio-economic models into a holistic system. Agricultural Systems, 75(1), 1-21.

[2] Kusiak, A. (2000). Data Analysis: Models and Algorithms. In: Proceedings of the SPIE Conference on Intelligent Systems and Advanced Manufacturing.(Edited by P.E. Orban and G.K. Knopf) , SPIE, Vol. 4191, Boston, MA, November 2000, pp. 1-9.

[3] Rai, A., Malhotra, P.K., Sharma, S.D. and Chaturvedi, K. K. Data Warehousing for Agricultural Research - An Integrated Approach for Decision Making. Journal of the Indian Society of Agricultural Statistics. 61(2): 264-273.

[4] Weibelzahl, S., Jedlitschka, A., & Ayari, B. (2006). Eliciting Requirements for an Adaptive Decision Support System through Structured User Interviews. In: Proceedings of the Fifth Workshop on User-Centred Design and Evaluation of Adaptive Systems, held in conjunction with the 4th International Conference on Adaptive Hypermedia & Adaptive Web-based Systems.(Edited by Weibelzah S. et al.) (AH'06), Dublin, Ireland, June 20th, 2006, pp. 770-778.

[5] Perini, A., Susi, A., & ITC-irst, T. P. I. (2011). Understanding the requirements of a decision support system for agriculture. An agent-oriented approach management, 11, 12.

[6] Inmon W.H. (1993) .Building the Data Warehouse. A Wiley QED publication, John Wiley and Sons, Inc. New York. 123pp.

[7] Giorgini.P, Rizzi.S, and Garzetti.M. (2005). Goal-oriented requirements analysis for data warehouse design. In: Proceedings of the 8th ACM International Workshop on Data Warehousing and OLAP, pp 47-60.

[8] Anton, A. I. (1996). Goal-based requirements analysis. In: Proceedings of the IEEE Second International Conference on Requirements Engineering, pp. 136-144.

[9] List, B., Bruckner, R., Machaczek, K. and Schiefer, J. (2002). A comparison of data warehouse development methodologies: Case study of the process warehouse. In: Proceedings of the DEXA, pp 203–215.

[10] Wong, B., Wong, K., Schneider, A., Maloney, G. (1999). Data Warehousing. Bentley College Community, pp 56-75.

[11] Strauss, A., & Corbin, J. M. (1990). Basics of qualitative research: Grounded theory procedures and techniques. Sage Publications, Inc.

[12] Brunette, W., Sundt, M., Dell, N., Chaudhri, R., Breit, N. and Borriello, G. (2013). Open Data Kit 2.0: Expanding and Refining Information Services for Developing Regions.

[13] Bonifati, A., Cattaneo, F., Ceri, S., Fuggetta, A. and Paraboschi, S (2001). Designing data marts for data warehouses. ACM Transactions on Software Engineering and Methodology. 10(4):452-483.

[14] Bruckner, R.M., List, B. and Schiefer, J. (2001).A Holistic Approach for Managing Requirements of Data Warehouse Systems. In: Proceedings of the Eighth Americas Conference on Information Systems.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[15] Bruckner, R.M., List, B. and Schiefer, J. (2001).Developing Requirements for Data Warehouse Systems with Use Cases. In: Proceedings of the Seventh Americas Conference on Information Systems.
[16] Poe, V., Brobst, S., and Klauer, P. (1997). Building a data warehouse for decision support. Prentice-Hall, Inc.
[17] Goguen J. A, Linde, C. (1993). Techniques for Requirement Elicitation. In: Proceedings of Conference on Requirement Engineering.
[18] Inmon, Bill (2005). Building the Data Warehouse. Fourth Edition, John Wiley, NewYork.
[19] Sarkar, A. (2012). Data Warehouse Requirements Analysis Framework: Business-Object Based Approach. *International Journal of Advanced Computer Science and Applications.* 3 (1), p25-34.
[20] Rai, A., Dubey, V., Chaturvedi, K. K., Malhotra, P. K. (2008). Design and development of data mart for animal resources. *Journal Computers and Electronics in Agriculture.* 64 (2), p111-119.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Computational Algorithms Based on the Paninian System to Process Euphonic Conjunctions for Word Searches

Rajitha V.
Department of Computer Science
Meenakshi College for Women
Chennai, India
&
Research Scholar
Mother Teresa Women's University
Kodaikanal, India

Kasmir Raja S. V.
Dean – Research
SRM University
Chennai, India

Meenakshi Lakshmanan
Department of Computer Science
Meenakshi College for Women
Chennai, India
meenakshi.lakshmanan@gmail.com

***Abstract* – Searching for words in Sanskrit E-text is a problem that is accompanied by complexities introduced by features of Sanskrit such as euphonic conjunctions or '*sandhis*'. A word could occur in an E-text in a transformed form owing to the operation of rules of *sandhi*. Simple word search would not yield these transformed forms of the word. Further, there is no search engine in the literature that can comprehensively search for words in Sanskrit E-texts taking euphonic conjunctions into account. This work presents an optimal binary representational schema for letters of the Sanskrit alphabet along with algorithms to efficiently process the *sandhi* rules of Sanskrit grammar. The work further presents an algorithm that uses the *sandhi* processing algorithm to perform a comprehensive word search on E-text.**

Keywords – Sanskrit; euphonic conjunction; sandhi; linguistics; Panini; Sanskrit word search; E-text search.

I. INTRODUCTION

Word search in Sanskrit E-texts is a problem that is beset with complexities, unlike in the case of English E-texts. The problem assumes relevance in the context of the availability of rapidly increasing numbers of ancient Sanskrit texts [5-9] in the electronic format. The importance of n-gram analysis of Sanskrit texts for scholars and the tremendous utility of locating specific words in a variety of texts to aid the scholastic process can hardly be overemphasized.

Dating of a text, fixing its authorship with certainty, and analysis of the writing style of an author of a text, are some of the areas in which n-grams assume criticality especially in the context of ancient Sanskrit works. Quoting from authoritative texts is imperative in scholarly works, and word searches can provide crucial help in this regard. Locating the portion in a text or texts in which a particular usage or word is found is of great importance to scholars who write explanatory treatises of Sanskrit-based works in English and other languages. Semantic analysis and understanding of texts are facilitated by finding occurrences of words and studying them in different contexts. In fact, ancient Sanskrit works are universally acknowledged as being mines of information on a whole spectrum of disciplines, and hence finding actual occurrences of words is of great consequence to not only Sanskrit scholars but to also researchers from various other disciplines ranging from philosophy, theology, the arts and the physical and life sciences to sociology, medicine and astronomy.

II. THE PROBLEM

As stated above, there are complexities involved in searching comprehensively for words in a Sanskrit E-text. One of the major contributors to this complexity is the operation of euphonic conjunctions or '*sandhis*'. A *sandhi* is a point in a word or between words, at which adjacent letters coalesce and transform [3]. This is a common feature in many Indian languages as against European languages, and has far-reaching consequences in Sanskrit. The transformation caused by the application of rules of *sandhi* in Sanskrit can be significant enough to alter the word itself to such a degree that the transformed word would not show up in a simple word search.

For example, the word '*asamardhiḥ*' (meaning of unmatched affluence), can be transformed into '*āsamardhiḥ*' because of the operation of a euphonic conjunction with a word ending in '*a*' preceding it, or '*āsamardhir*' or '*āsamardhis*' in combination with words occurring after it or '*asamarddhiḥ*' or '*asamardddhiḥ*' by internal transformation. Clearly, simply searching for the word *asamardhiḥ* would not yield the occurrences of the same word as '*asamarddhir*', '*āsamardddhir*', or other alternative forms. As such, a normal text-search using a Unicode text editor would not suffice. Other search engines currently used for Sanskrit [13] too do not provide for such comprehensive searching.

In order to achieve such an exhaustive search, all possible forms of the word resulting from the euphonic conjunctions that would become operative in its case must be generated and searched for in the given text.

The authors have already presented a new schema for fast *sandhi* processing in earlier work [4]. The present work extends the application of that schema to other *sandhi* rules including consonant-based and *visarga*-based *sandhis* as well as important rules with respect to exceptional cases. It further presents a

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

complete computational algorithm to process all *sandhis*, and an algorithm to apply this *sandhi*-processing procedure to generate all word forms to enable comprehensive searching.

A. Language Representation

The Unicode hexadecimal range 0900 - 097F is used to represent Sanskrit characters in *Devanāgarī* script. The characters used to represent Sanskrit letters in English script are found in the Basic Latin (0000-007F), Latin-1 Supplement (0080-00FF), Latin Extended-A (0100-017F) and Latin Extended Additional (1E00 – 1EFF) Unicode ranges.

The Latin character set has been employed in this work to represent Sanskrit letters as E-text. As such, the schema and algorithms presented do not use *Devanāgarī* script. To use the algorithms for text that is in *Devanāgarī* script, the text needs to first be converted to Latin text.

B. Terminology

The terminology employed in this work for certain groups of letters of the Sanskrit alphabet is given in Table 1.

Table 1: Terminology

Term	Description / Notation
Vowel	*a,ā,i,ī,u,ū,ṛ,ṝ,ḷ,e,ai,o,au*
Semi-vowel	*y, v, r, l*
Consonant	*k,kh,g,gh,ṅ,c,ch,j,jh,ñ,ṭ,ṭh,ḍ,ḍh,ṇ,t,th,d,dh, n,p,ph,b,bh,m,ś,ṣ,s,h*
Guttural	*k, kh, g, gh, ṅ*
Palatal	*c, ch, j, jh, ñ*
Cerebral	*ṭ, ṭh, ḍ, ḍh, ṇ*
Dental	*t, th, d, dh, n*
Labial	*p, ph, b, bh, m*
Nasal	*ṅ, ñ, ṇ, n, m*
Aspirate	*h*
Sibilant	*ś, ṣ, s*
Column1	*k, c, ṭ, t, p*
Column2	*kh, ch, ṭh, th, ph*
Column3	*g, j, ḍ, d, b*
Column4	*gh, jh, ḍh, d, bh*
Visarga	*ḥ*
Anusvāra	*ṁ*
Hard consonant	*Column1, Column2, Sibilants*
Soft consonant	*Column3, Column4, Nasals, Aspirate*
Hard guttural	*k, kh*
Hard labial	*p, ph*
Mutes	*Column1, Column2, Column3, Column4, Nasals*
Jihvāmūlīya	ʎ (pronounced as the end of '*kah*')
Upadhmānīya	ϒ(pronounced as the end of '*paf*')

III. The Basis of the Work

The renowned ancient Sanskrit linguist, Pāṇini, codified the extant grammar of Sanskrit into terse aphorisms ('*sūtras*') and organized these aphorisms into eight chapters. This work is the authoritative *Aṣṭādhyāyī* (literally meaning 'work in eight chapters') and is universally acknowledged as the most comprehensive codification of the grammar of any language. The grammatical rules that make up Pāṇini's *Aṣṭādhyāyī* are derivational and known for their mathematical precision in spite of dealing with the nuances of the language at various levels including morphology, syntax, semantics, phonology and pragmatics. Owing to the cryptic nature of the *Aṣṭādhyāyī*, one or more of the commentaries on it are required to get a clear understanding of its contents.

The current work deals with Pāṇini's *sandhi*-related aphorisms with the help of the recognized commentaries, *Siddhānta-kaumudī* [1] and *Kāśikā* [2]. Both these commentaries are accepted by Sanskrit scholars as authoritative works on Pāṇinian grammar.

Pāṇini's statements of grammatical rules are expressed on the basis of the *Māheśvara-sūtras*, or the 'aphorisms of Maheśvara'. These aphorisms provide a list of all the letters of the Sanskrit alphabet ordered in a specific sequence. The *Māheśvara* aphorisms are given below:

1. *a-i-u-ṇ*
2. *ṛ-ḷ-k*
3. *e-o-ṅ*
4. *ai-au-c*
5. *ha-ya-va-ra-ṭ*
6. *la-ṇ*
7. *ña-ma-ṅa-ṇa-na-m*
8. *jha-bha-ñ*
9. *gha-ḍha-dha-ṣ*
10. *ja-ba-ga-ḍa-da-ś*
11. *kha-pha-cha-ṭha-tha-ca-ṭa-ta-v*
12. *ka-pa-y*
13. *śa-ṣa-sa-r*
14. *ha-l*

The last letter in each of these aphorisms is only a place-holder. The first four aphorisms list only the short forms of all the vowels, while the rest list the semi-vowels and consonants; the latter list has the vowel '*a*' appended to each letter only to enable pronunciation of the aphorism.

A. The Approach

The present work is based on earlier work by the authors, which directly codifies Pāṇini's rules in a novel way using binary representations [4]. The unique data representation devised by the authors has been further refined in this work and consonant-based, *visarga*-based *sandhi* rules, as well as some special *sandhi* rules have been included in this work.

Rule representation has been simplified to minimal binary set-unset operations. Further, the *sūtra* ordering has been done after acquiring a thorough understanding of the operation of Pāṇini's *sandhi*-related aphorisms. As such, this work presents a significant extension, refinement and closure of the earlier work of the authors. Moreover, it provides a clear understanding of the rules governing *sandhi* as laid down by Pāṇini, in a comprehensive and simplified way, hitherto not encountered in the literature.

B. The Binary Schema

The following is an extract from already published work by the authors [4] and is included here for completeness of the presentation.

A point of *sandhi* is denoted by

$$x + y$$

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

where x and y denote the *sandhi* letters and the symbol '+' denotes adjacency. The variable X denotes the sequence of letters culminating in x; the variable Y denotes the sequence of letters starting with y. The notations X and Y are used to depict special conditions that pertain to an entire word or sequence of letters involved in the *sandhi* rule. The letter immediately preceding x and the letter immediately succeeding y are denoted respectively as u and w respectively.

The refined schematic developed in this work to represent letters of the Sanskrit alphabet is given in Table 2.

Table 2: Binary representation scheme

#	**Letters**
0	*a,ā,i,ī,u,ū,ṛ,ṝ,ḷ,e,ai,o,au*
1	*y,r,l,v,yaṁ,vaṁ,laṁ*
2	*k,kh,g,gh,ṅ,c,ch,j,jh,ñ,ṭ,ṭh,ḍ,ḍh,ṇ,t,th,d,dh,n,p,ph,b,bh,m,ś,ṣ,s,h*
3	*ṁ,ḥ,',#,x,f*
4	*a*
5	*ā*
6	*u,i*
7	*ū,ī*
8	*u,i,ṛ,ḷ,a*
9	*ū,ī,ṝ,ṝ,ā*
10	*u,i,ṛ,ḷ*
11	*ū,ī,ṝ*
12	*ṛ,ṝ,ḷ*
13	*o,e*
14	*au,ai*
15	*o,au,e,ai*
16	*o,e,ar,al*
17	*ār,ār,āl*
18	*av,āv,ay,āy*
19	*ava*
20	*v,y,r,l*
21	*r*
22	*ṁv,ṁy,r,ṁl*
23	*ṣ*
24	*s*
25	*ś*
26	*h*
27	*ṅ,ṇ,n,ñ,m*
28	*ṅ,ṇ*
29	*n*
30	*m*
31	*k,gh,kh,g,ṅ*
32	*c,jh,ch,j,ñ,ś*
33	*ṭ,ḍh,ṭh,ḍ,ṇ,ṣ*
34	*t,dh,th,d,n,s*
35	*p,bh,ph,b,m*
36	*k,ṭ,t,c,p*
37	*kh,ṭh,th,ch,ph*
38	*g,ḍ,d,j,b*
39	*gh,ḍh,dh,jh,bh*
40	*ch,ṭh,th,c,ṭ,t*
41	*k,kh,p,ph*
42	*x,x,f,f*
43	*ṁ*
44	*ḥ*
45	*'*
46	*#*

In Table 2, *x* is the *jihvāmūliyā* and *f* is the *upadhmānīya* mentioned in Table 1; the # symbol stands for '*ru*' a special intermediary form of the semi-vowel *r*.

Any letter of the alphabet is represented in two parts: Part 1 denotes the category to which a letter belongs (zero-based serial number in Table 2), and Part 2 denotes the zero-based term number within the series that the letter is or fits into. In any letter representation, Part 1 is a binary string of fixed length 46, in which the set bit denotes the category number, while Part 2 is a binary string of maximum length 6 in which the set bit indicates which particular letter is being represented. It is clear that one letter has many representations under this scheme.

The first four shaded rows of Table 2 stand for overall categories, viz. vowels, semi-vowels, consonants and special characters respectively. One of these four bits have to be set in any letter representation. There is no corresponding Part 2 value for the bits 0, 1, 2 and 3 of Part 1.

For simplicity of presentation, *sandhi* rules use the following notation: $x_i(n) = 1$ indicates that the nth bit of Part i of the variable x is set, where $i = 1, 2$. In the implementation, the checks for bit set can be done by simply using the XOR operation.

IV. *SANDHI* PROCESSING UNDER THE PĀṆINIAN SYSTEM

Each of the eight chapters of Pāṇini's *Aṣṭādhyāyī* is divided into four parts or *pādas*. Overall, the work is defined by Pāṇini as consisting of two parts, the *sapādasaptādhyāyī* (the aphorisms of Chapters 1.1 to 8.1), and the *tripādī* (the aphorisms of Chapters 8.2 to 8.4).

The order in which the rules should be visited was arrived at in this work after a thorough study of Pāṇini's aphorisms with respect to euphonic conjunctions. As a result of the study, the set of *sandhi* rules has been split into two in this work: Set 1, having all the relevant aphorisms of the *sapādasaptādhyāyī* as well as a few specific rules from the *tripādī*, and Set 2, having all the remaining relevant aphorisms of the *tripādī*.

The order of parsing is as follows: rules in Set 1 are parsed in reverse order of their *Aṣṭādhyāyī* order; rules in Set 2 are parsed in the *Aṣṭādhyāyī* order itself. (The *sūtra* number as it appears in the *Aṣṭādhyāyī* is indicated in the algorithm between double pipe symbols given after the *sūtra*.)This parsing order is adopted so that no rule already parsed has to be parsed again. As such, the flow of the program is just from top to bottom. There are exceptions to the above parsing orders in both sets that arise because of certain overruling *sūtras* that appear earlier / later respectively in the two sets. The ordering is changed to accommodate such rules in such a way as to parse them before the main rule.

Assuming that the rules are ordered in the above manner in the two sets, the following general algorithm for parsing rules is presented. The *word_list* is a list of the alternative word-pair outputs generated, and represents the output at the end of the algorithm.

Algorithm SandhiRulesParser

```
{
    while Set 1 rules have not been fully visited
    {
        Try the next rule in Set 1;
        if the rule applies
        {
            Apply the rule and store the output;
            if the rule is optional
            {
                Add the current word pair to the word_list;
                Continue checking from the next rule for all
                word pairs stored in the word_list;
            }
            else
            {
                Process internal sandhi rules of Set 2;
            }
        }
    }
    while Set 2 rules have not been fully visited
    {
        Try the next rule in Set 2;
        if the rule applies
        {
            Apply the rule and transform the given words;
            if the rule is optional
            {
                Add the current word pair to the word_list;
                Continue checking from the next rule for all
                word pairs stored in the word_list;
            }
        }
    }
}
```

There are a few exceptions that would apply to the general processing order prescribed by the above algorithm. For example, in Set 1 the output produced by an optionally applying rule does, in certain cases, have to pass through a rule appearing below and undergo further transformation as a result, as it happens in Set 2. Also, it is found that a few rare rules of Set 2 have to be processed before Set 1. Further, in Set 2, all rules that form exceptions to a particular rule are stated after it by Pāṇini, but clearly, have to be processed before the rule by the algorithm.

V. The Sandhi Processor

The key to symbols used in rule coding and algorithm specification is as follows:

- // means single-line explanatory comment
- { } are block or set indicators
- ∧ denotes and
- ∨ denotes or
- ¬ denotes not
- ⊕ denotes xor
- | denotes word concatenation

The algorithm SandhiProcessor processes the rules pertaining to all the major *sandhis* in Sanskrit grammar, in accordance with the processing scheme presented in Algorithm SandhiRulesParser. Set 1 and Set2 *sandhis* have been incorporated here one below the other and the rules have been codified as per the schema presented above. The vowel *sandhis* presented in [4] have been modified in accordance to the reduced schema and included here for completeness.

When x_i, y_i, etc., for i = 1,2, are assigned new values by setting bits, it is assumed that their initial values are first unset. Also, if either part of a variable is not set, it is assumed to remain unchanged. Further, it is also assumed that a category change caused by a *sandhi* will cause an automatic change in the first four bits of the letter representation and that all bit representations for the changed letter will be generated thereafter. Hence, these aspects are not explicitly stated for each rule in this algorithm.

The speed of processing is increased by going into a rule only if overall conditions are satisfied. For instance, if the rule is a vowel *sandhi* rule, where both *x* and *y* are required to be vowels, then the check if $x_1(0)$ and $y_1(0)$ are 1 is first made. If this bit-check is not true for the input words, then a whole set of vowel *sandhis* is omitted from the parse, thus increasing the efficiency of the algorithm. These overall checks have not been shown in the algorithm presented below, in order to make the presentation more simple.

Algorithm SandhiProcessor (X,Y)

```
{
    //1.svaujasamauṭchaṣṭābhyāṁbhisṅebhyāṁbhyasṅasi
    //bhyāṁbhyasṅasosāmṅyossup || 4.1.2 ||
    //If there is a visarga (ḥ) at the end of X, then the visarga is
    //changed to 's'.
    if x1(44)
    {
        x1(24) = 1;
    }

    //2. sasajuṣo ruḥ || 8.2.66 ||
    // Common name: visarga-rutva sandhi
    //If last letter of X is s, then s is replaced by '#' which
    //stands for the particle 'ru', interpreted as 'r'.
    //This rule is incorporated here though it belongs to the
    //sapādasaptādhyāyī .
    if x1(24)
    {
        x1(46) = 1;
    }

    //3. avaṅ sphoṭāyanasya || 6.1.123 || (vowel sandhi)
    // Common name: avaṅādeśa sandhi
    //If the word go is followed by a vowel, then the o of go
    //is optionally replaced by ava.
    if (x1(15) ∧ x2(0)) ∧ (u1(31) ∧ u2(3))
    {
        Add X|Y to word_list;
        x1 (19) = 1;
    }
    //4. haśi ca || 6.1.114 ||
    //If # (ru) or r at the end of X is preceded by the vowel a
    //and followed by aspirate, semi-vowel, nasal, Column3 or
```


http://sites.google.com/site/ijcsis/
ISSN 1947-5500

//Column4, then last letter of *X* is replaced with the vowel
//'u' and shifted to Y to become the first letter of Y
$if\ (x_1(46) \vee x_1(21)) \wedge u_1(4) \wedge (y_1(1) \vee y_1(26) \vee y_1(27) \vee y_1(38) \vee y_1(39))$
{
    $x_1(6) = 1;$
    $x_2 = u_2;$
    Shift x from the end of *X* to the beginning of *Y*;
}

//5. *ato roraplutādaplute* || 6.1.113 ||
//Common name: *visarga-rutva sandhi*
//If # (*ru*) or *r* at the end of *X* is followed and preceded by
//*a*, then the # or *r* is replaced with the vowel '*u*' and shifted
//to Y to become the first letter of Y
$if\ (x_1(46) \vee x_1(21)) \wedge u_1(4) \wedge y_1(4)$
{
    $x_1(6) = 1;$
    $x_2 = u_2;$
    Shift x from the end of *X* to the beginning of *Y*;
}

//6. *eṅaḥ padāntādati* || 6.1.109 || (vowel *sandhi*)
// Common name: *pūrvarūpa sandhi*
//If *e* or *o* at the end of a word is followed by *a*, then *e* or *o*
//remains, and the *avagraha* (') replaces *a*.
$if\ x_1(13) \wedge\ y_1(4)$
{
    $y_1(45) = 1;$
}

//7. *akaḥ savarṇe dīrghaḥ* || 6.1.101 || (vowel *sandhi*)
//Common name: *savarṇadīrgha sandhi*
//If one of *a, i, u, ṛ* or *ḷ* or their long equivalents *ā, ī, ū* and *ṝ*
//is followed by the short or long form of the same letter,
//then the corresponding long letter replaces both.
$if\ (x_1(8) \vee x_1(9)) \wedge\ (y_1(8) \vee y_1(9)) \wedge \neg(x_2 \oplus y_2)$
{
    delete *y*;
    $x_1(9) =\ 1;$
    return $X|Y$;
}

//8. *omāṅośca* || 6.1.95 || (vowel *sandhi*)
// Common name: *pararūpa sandhi*
//If *a* or *ā* is followed by *o* of the word *om* or *oṁ*, then *o*
//replaces both.
$if(x_1(4) \vee x_1(5)) \wedge Y \in \{om, oṁ\}$
{
    delete x;
}

//9. *etyedhatyūṭhsu* || 6.1.89 || (vowel *sandhi*)
//Common name: *vṛddhi sandhi*
//For this rule, in all cases the resultant letter replaces *x*
//and *y*.
//i) If *a* or *ā* is followed by *eti* or *edhati*, then *vṛddhi* letter
//*ai* replaces both
//ii) If the preposition *pra* is followed by *eṣa* or *eṣy*, then
//*vṛddhi* letter *ai* replaces both
//iii) If *a* or *ā* is followed by *ūh*, then *vṛddhi* letter *au*
//replaces both
//iv) If preposition *pra* is followed by *ūḍh*, then *vṛddhi*
//letter *au* replaces both
//v) If word *sva* is followed by *īr*, then *vṛddhi* letter *ai*
//replaces both
$if\ (x_1(4) \vee x_1(5))$ //x is '*a*' or '*ā*'
{
    $if\ (y_1(13) \wedge y_2(1))$ //y is '*e*'
    {
        $if\ Y$ starts with {*et, edhat*} //(i)
        {
            delete x;
            $y_1(14) = 1;$
        }
        $elseif\ X$ = '*pra*' $\wedge\ Y$ starts with {*eṣ, eṣy*} //(ii)
        {
            delete x;
            $y_1(14) = 1;$
        }
    }
    $elseif\ (y_1(7) \wedge y_2(0))$ //y is '*ū*'
    {
        $if\ w_1(26)$ // (iii)
        {
            delete x;
            $y_1(14) = 1;$
        }
        $elseif\ X$ = '*pra*' $\wedge\ Y$ starts with {*ūḍh*} //(iv)
        {
            delete x;
            $y_1(14) = 1;$
        }
    }
    $elseif\ X$ = '*sva*' $\wedge (y_1(7) \wedge y_2(1)) \wedge w_1(21)$//(v)
    {
        delete x;
        $y_1(14) = 1;$
    }
}

//10. *eṅi pararūpaṁ* || 6.1.94 || (vowel *sandhi*)
// Common name: *pararūpa sandhi*
//If *a* or *ā* at the end of a preposition is followed by *e* or
//*o*, then the *e* or *o* replaces both.
//Note: The prepositions that qualify are: *pra, ava, apa,*
//upa, parā.
$if\ X \in \{pra, ava, apa, upa, parā\} \wedge y_1(13)$
{
    delete x;
}

//11. *upasargādṛti dhātau* || 6.1.91 || (vowel *sandhi*)
//Common name: *vṛddhi sandhi*
//i) If *a* or *ā* at the end of a preposition is followed by *ṛ*,
//ṝ or *ḷ*, then *vṛddhi* letter *ār, ār* or *āl* respectively
//replaces both. *Note:* The prepositions that qualify are:
//*pra, parā, apa, ava, upa*

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

//ii) If the word *vatsara, kambala, vasana, daśa, ṛṇa* is
//followed by the word *ṛṇa*, then *vṛddhi* letter *ār*
//replaces both.
//Note: This rule clashes with 6.1.87 (*guṇa sandhi*), and
//takes precedence.
$if\ X\ \in$ {*pra, ava, apa, upa, parā*} $\wedge\ y_1(12)$
{
delete x;
$y_1(17) = 1$;
}
$if\ X\ \in$ {*vatsara, kambala, vasana, daśa, ṛṇa*} $\wedge$
$(y_1(12) \wedge y_2(0)) \wedge Y =$ '*ṛṇa*'
{
delete x;
$y_1(17) = 1$;
}

//12. *vṛddhireci* || 6.1.88 || (vowel *sandhi*)
// Common name: *vṛddhi sandhi*
//If *a* or *ā* is followed by *e, o, ai* or *au*, then the
//corresponding *vṛddhi* letter *ai* or *au* replaces both.
$if\ (x_1(4) \vee x_1(5)) \wedge ((y_1(13) \vee y_1(14))$
{
delete x;
$y_1(14) = 1$;
}

//13. *ādguṇaḥ* || 6.1.87 || (vowel *sandhi*)
// *uraṇ raparaḥ* || 1.1.51 ||
// Common name: *guṇa sandhi*
//If *a* or *ā* is followed by *i, ī, u, ū, ṛ, ṝ* or *ḷ*, then the
//corresponding *guṇa* letter *e, o, ar* or *al* replaces both.
$if\ (x_1(4) \vee x_1(5)) \wedge ((y_1(10) \vee y_1(11))$
{
delete x;
$y_1(16) = 1$;
}

//14. *ecoyavāyāvaḥ* || 6.1.78 || (vowel *sandhi*)
// Common name: *ayāyāvāvādeśa sandhi*
//If *e, o, ai* or *au* is followed by a vowel, then *ay, av, āy,*
//āv replace the first respectively.
$if\ x_1(15) \wedge y_1(0)$
{
$x_1(18) = 1$;
}

//15. *iko yaṇaci* || 6.1.77 || (vowel *sandhi*)
// Common name: *yaṇādeśa sandhi*
//If *i, ī, u, ū, ṛ, ṝ* or *ḷ* is followed by a vowel, then the
//corresponding semi-vowel (*y, v, r, l*) replaces the first.
$if\ (x_1(10) \vee x_1(11)) \wedge y_1(0)$
{
$x_1(20) = 1$;
}

//16. *che ca* || 6.1.73 ||
//Common name: *tugāgama sandhi*
//If a short vowel is followed by the consonant *ch*, then
//*t* is added.
$if\ x_1(8) \wedge (y_1(40) \wedge y_2(0))$
{
$z_1(34) = 1$;
$z_2 = y_2$;
Add z to the end of X;
}

//17. *āṅmāṅośca* || 6.1.74 ||
//Common name: *tugāgama sandhi*
//If the particle *ā* or word *mā* is followed by *ch*,
//then *t* is added.
$if\ X \in$ {*ā, mā*} $\wedge$ ($y_1(40) \wedge y_2(0)$)
{
$z_1(34) = 1$;
$z_2 = y_2$;
Add z to the end of X;
}

//18. *dīrghāt* || 6.1.75 ||
// *padāntādvā* || 6.1.76 ||
//Common name: *tugāgama sandhi*
//If a long vowel is followed by *ch*, then *t* is added.
$if\ x_1(9) \wedge (y_1(40) \wedge y_2(0))$
{
$z_1(34) = 1$;
$z_2 = y_2$;
Add z to the end of X;
}

//19. *saṁyogāntasya lopaḥ* || 8.2.23 ||
//If the final consonant of X is preceded by a
//consonant, then the last consonant is dropped.
$if\ x_1(2) \wedge u_1(2)$
{
delete x;
}

//20. *jhalām jaśo'nte* || 8.2.39 ||
//Common name: *jaśtva sandhi*
//If x is Column1, Column2, Column3, Column4, sibilant
//or aspirate, then x is replaced by the corresponding
//Column3.
//Note: The rule *sasajuṣo ruḥ* || 8.2.66 || debars the
//application of this rule for words ending in sibilants, and
//has been incorporated earlier in Set 1 rules itself.
$if\ x_1(36) \vee x_1(37) \vee x_1(38) \vee x_1(39)$
{
$x_1(38) = 1$;
}

//21. *pumaḥ khayyampare* ||8.3.6||
// *atrānunāsikaḥ pūrvasya tu vā* || 8.3.2 ||
//If X is the word '*pum*' or '*puṁ*' and is followed by
//Column1 or Column2, which is in turn followed by a
//vowel, semi-vowel or a nasal, then x is replaced by #
//(*ru*) and the preceding vowel is made nasal using the

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

//*anusvāra*.
$if\ X \in \{pum, puṁ\} \wedge (y_1(36) \vee y_1(37)) \wedge (w_1(0) \vee w_1(1) \vee w_1(27))$
{
$x_1(43) = 1;$
$z_1(46) = 1;$
$x_2(0) = z_2(0) = 1;$
}

//22. *naśchavyapraśān* || 8.3.7 ||
// *atrānunāsikaḥ pūrvasya tu vā* || 8.3.2 ||
//Common name: *satva sandhi*
//If the final *n* of a word, except for the word *praśān*,
//followed by *ch, ṭh, th, c, ṭ* or *t* which is in turn
//followed by a vowel, semi-vowel or nasal, then
//*n* is replaced with # (*ru*) and the preceding vowel is
//made nasal using the *anusvāra*.
$if\ \neg(X = \text{'}praśān\text{'}) \wedge (x_1(27) \wedge x_2(2)) \wedge y_1(40)) \wedge (w_1(0) \vee w_1(1) \vee w_1(27))$
{
$u = x;$
$x_1(43) = 1;$
$z_1(46) = 1;$
$x_2(0) = z_2(0) = 1;$
}

//23. *nm̐pe* || 8.3.10 ||
// *atrānunāsikaḥ pūrvasya tu vā* || 8.3.2 ||
//If X = '*nm̐n*' and y is the letter '*p*' then x is
//optionally replaced with # (*ru*) and the preceding
//vowel is made nasal using the *anusvāra*.
$if\ X = \text{'}nm̐n\text{'} \wedge (y_1(35) \wedge y_2(0))$
{
Add *X*|*Y* to *word_list*;
$u = x;$
$x_1(43) = 1;$
$z_1(46) = 1;$
$x_2(0) = z_2(0) = 1;$
}

//24. *ro ri* || 8.3.14 ||
// *ḍhralope pūrvasya dīrgho'ṇaḥ* ||6.3.111||
//If *r* or # (*ru*) is followed by *r* and is preceded by *a, i* or *u*,
//then one *r* is dropped and the short vowel is made long.
$if\ (x_1(46) \vee x_1(21)) \wedge y_1(21)$
{
delete x;
$if\ u_1(6)$
{
$u_1(7) = 1;$
}
$if\ u_1(4)$
{
$u_1(5) = 1;$
}
}

//25. *kharavasānayorvisarjanīyaḥ* || 8.3.15 ||
//If x is # (*ru*) or *r* and is followed by a hard consonant,
//then x is replaced with *visarga*.
$if\ (x_1(46) \vee x_1(21)) \wedge (y_1(36) \vee y_1(37) \vee y_1(23) \vee y_1(24) \vee y_1(25))$
{
$x_1(44) = 1;$
}

//26. *bhobhagoaghoapūrvasya yo'śi* || 8.3.17 ||
//If x is # (*ru*) or *r* and is preceded by *bho, bhago, agho,*
//*a* or *ā*, and is followed by a vowel, semi-vowel or soft
//consonant, then x is replaced by the consonant '*y*'.
$if\ (x_1(46) \vee x_1(21)) \wedge (X - \{x\})\ in\ \{bho, bhago, agho, a, ā\} \wedge (y_1(0) \vee y_1(1) \vee y_1(38) \vee y_1(39) \vee y_1(26) \vee y_1(27))$
{
$x_1(20) = 1;$
$x_2(1) = 1;$

}

//27. *lopaḥ śākalyasya* || 8.3.19 ||
//If the consonant '*y*' or '*v*' is preceded by *a* or *ā* and is
//followed by a vowel, semi-vowel or soft consonant,
//then the '*y*' or '*v*' is dropped.
$if\ (x_1(20) \wedge (x_2(0) \vee x_2(1))) \wedge (u_1(4) \vee u_1(5)) \wedge (y_1(0) \vee y_1(1) \vee y_1(38) \vee y_1(39) \vee y_1(26) \vee y_1(27))$
{
delete x;
}

//28. *oto gārgyasya* || 8.3.20 ||
//If the consonant '*y*' is preceded by the vowel '*o*' and
//followed by a vowel, semi-vowel or soft consonant,
//then the '*y*' is dropped.
$if\ (x_1(20) \wedge x_2(1)) \wedge (u_1(13) \wedge u_2(0)) \wedge (y_1(0) \vee y_1(1) \vee y_1(38) \vee y_1(39) \vee y_1(26) \vee y_1(27))$
{
delete x;
}

//29. *uñi ca pade* || 8.3.21 ||
//If the consonant '*y*' or '*v*' is preceded by *a* or *ā* and is
//followed by the word '*u*', then the '*y*' or '*v*' is dropped.
$if\ (x_1(20) \wedge (x_2(0) \vee x_2(1))) \wedge (u_1(4) \vee u_1(5)) \wedge Y =$ *'u'*
{
delete x;
}

//30. *hali sarveṣāṁ* || 8.3.22 ||
//If the consonant '*y*' is followed by a semi-vowel or
//consonant, then the '*y*' is dropped.
$if\ (x_1(20) \wedge x_2(1)) \wedge (y_1(1) \vee y_1(2))$
{
delete x;

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

}

//31. *he mapare vā* || 8.3.26 ||
//i) If *m* is followed by *h* at the end of a word which is in
//turn followed by *m*, then the first *m* is optionally changed
//to *anusvāra*.
//ii) If *m* is followed by *h* which is in turn followed by
//consonants '*y*', '*l*', or '*v*', then the *m* is optionally replaced
//by the nasal forms of '*y*', '*l*', or '*v*' respectively.
$if\ x_1(30) \wedge y_1(26)$
{
    $if\ w_1(30)$
    {
        Add *X*|*Y* to *word_list*;
        $x_1(43) = 1;$
    }
    $elseif\ w_1(20)$
    {
        $x_1(22) = 1;$
        $x_2 = w_2;$
    }
}

//32. *napare naḥ* || 8.3.27 ||
//If *m* is followed by *h* at the end of a word which is in turn
//followed by *n*, then the *m* is optionally replaced by *n*.
$if\ x_1(30) \wedge y_1(26) \wedge (w_1(27) \wedge w_2(2))$
{
    Add *X*|*Y* to *word_list*;
    $x = w;$
}

//33. *naścāpadāntasya jhali* || 8.3.24 ||
//If *n* is followed by Column1, Column2, Column3,
//Column4, sibilant or aspirate not at the end of a word,
//then the *n* is replaced by *anusvāra*.
for (each letter x in a word and its succeeding letter y)
{
    $if\ (x_1(29)) \wedge (y_1(36) \vee y_1(37) \vee y_1(38) \vee y_1(39) \vee y_1(23) \vee y_1(24) \vee y_1(25) \vee y_1(26))$
    {
        $x_1(43) = 1;$
    }
}

//34. *mo rāji samaḥ kvau* || 8.3.25 ||
//If *m* of the word '*sam*' or '*sām*' is followed by a word
//starting with '*rāj*' or '*rāṭ*' or '*ṛāñ*', then the *m* remains
//unchanged.
$if\ X \in \{sam, sām\} \wedge Y \in \{rāj, rāṭ, ṛāñ\}$
{
    //Skip Rule 35
    Continue processing from Rule 36;
}
//35. *mo'nusvāraḥ* || 8.3.23 ||
//Common Name*: anusvāra sandhi*
//If *m* at the end of a word is followed by any consonant,
//then *m* is replaced by *anusvāra*.
$if\ x_1(30) \wedge y_1(2)$
{
    $x_1(43) = 1;$
}

//36. *ṅaṇoḥ kuk ṭuk śari* || 8.3.28 ||
//If *ṅ* or *ṇ* is followed by a sibilant, then *k* or *ṭ* is optionally
//added respectively.
$if\ x_1(28) \wedge (y_1(23) \vee y_1(24) \vee y_1(25))$
{
    Add *X*|*Y* to *word_list*;
    $z_1(36) = 1;$
    $z_2 = x_2;$
    Add *z* to the end of *X*;
}

//37. *ḍaḥ si dhuṭ* || 8.3.29 ||
//Common name: *dhuḍāgama sandhi*
//If *ḍ* is followed by *s*, then *dh* is added optionally
$if\ (x_1(38) \wedge x_2(1)) \wedge y_1(24)$
{
    Add *X*|*Y* to *word_list*;
    $z_1(34) = 1;$
    $z_2 = x_2;$
    Add *z* to the end of *X*;
}

//38. *naśca* || 8.3.30 ||
//Common name: *dhuḍāgama sandhi*
//If *n* is followed by *s*, then *dh* is added optionally.
$if\ (x_1(27) \wedge x_2(2)) \wedge y_1(24)$
{
    Add *X*|*Y* to *word_list*;
    $z_1(39) = 1;$
    $z_2 = x_2;$
    Add *z* to the end of *X*;
}

//39. *śi tuk* || 8.3.31 ||
//Common name: *tugāgama sandhi*
//If *n* is followed by *ś*, then *t* is optionally added.
$if\ (x_1(27) \wedge x_2(2)) \wedge y_1(25)$
{
    Add *X*|*Y* to *word_list*;
    $z_1(36) = 1;$
    $z_2 = x_2;$
    Add *z* to the end of *X*;
}

//40. *ṅamo hrasvādaci ṅamuṇnityaṁ* ||8.3.32||
//Common name: *ṅamuḍāgama sandhi*
//If *ṅ*, *ṇ* or *n* is preceded by a short vowel and succeeded by
//a vowel, then the *ṅ*, *ṇ* or *n* gets duplicated.
$if\ (x_1(27) \wedge (x_2(0) \vee x_2(1) \vee x_2(2))) \wedge u_1(8) \wedge y_1(0)$
{
    Add another x to the end of *X*;

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

}

//41. *śarpare visarjanīyaḥ* || 8.3.35 ||
//If *visarga* is followed by a hard consonant which is in turn
//followed by a sibilant, then the *visarga* is retained.
$if\ x_1(44) \wedge (y_1(36) \vee y_1(37) \vee y_1(23) \vee y_1(24) \vee y_1(25)) \wedge (w_1(23) \vee w_1(24) \vee w_1(25))$
{
Store result with no change
}

//42. *vā śari* || 8.3.36 ||
//If *visarga* is followed by a sibilant, then the *visarga* is
//optionally retained.
$if\ x_1(44) \wedge (y_1(23) \vee y_1(24) \vee y_1(25))$
{
Add *X*|Y to *word_list*;
delete *x*;
}

//43. *kupvoḥ ʎk ϒpau ca* || 8.3.37 ||
//If *visarga* is followed by a hard guttural or hard labial,
//then it is replaced optionally by ʎ (pronounced as at the
//end of '*kah*') or ϒ (pronounced as at the end of '*paf*')
$if\ x_1(44) \wedge y_1(41)$
{
Add *X*|*Y* to word_list;
$x_1(42) = 1;$
$x_2 = y_2;$
Flag_kupvoh_sutra_fired = true //flag is set
}

//44. *so'padādau* || 8.3.38 ||
//If *visarga* is followed by *pāśa, kalpa, ka* or *kāmya*, then
//the *visarga* is replaced by *s*.
$if\ x_1(44) \wedge$ *Y* begins with {*pāśa, kalpa, ka, kāmya*}
{
$x_1(24) = 1;$
}

//45. *iṇaḥ ṣaḥ* || 8.3.39 ||
//If *visarga* is preceded by '*i*' or '*u*' and followed by *pāśa,
//kalpa, ka* or *kāmya*, then the *visarga* is replaced by *ṣ*.
$if\ x_1(44) \wedge u_1(6) \wedge$ *Y* begins with {*pāśa, kalpa, ka, kāmya*}
{
$x_1(23) = 1;$
}

//46. *namaspurasorgatyoḥ* || 8.3.40 ||
//If '*namaḥ*' or '*puraḥ*' is followed by a hard guttural or
//hard labial, then the *visarga* is replaced by *s* optionally.
$if\ X \in \{namaḥ, puraḥ\} \wedge y_1(41)$
{
Add *X*|*Y* to *word_list*;
$x_1(24) = 1;$
}

//47. *idudupadhasya cā'pratyayasya* || 8.3.41||
//If *visarga* is preceded by '*i*' or '*u*' and is at the end of any
//of *niḥ, duḥ, bahiḥ, āviḥ, catuḥ, prāduḥ* and is followed by
//a hard guttural or hard labial, then the *visarga* is replaced
//by *ṣ*.
$if\ X \in \{niḥ, duḥ, bahiḥ, āviḥ, catuḥ, prāduḥ\} \wedge y_1(41)$
{
$x_1(23) = 1;$
}

//48. *tiraso'nyatarasyām* || 8.3.42 ||
//If the word '*tiraḥ*' is followed by a hard guttural or hard
//labial, then the *visarga* is optionally replaced by *s*.
$if\ X =$ *'tiraḥ'* $\wedge y_1(41)$
{
Add *X*|Y to the word_list;
$x_1(24) = 1;$
}

//49. *dvistriścaturiti kṛtvorthe* || 8.3.43 ||
//If the words *dviḥ, triḥ* or *catuḥ* are followed by a hard
//guttural or hard labial, then the *visarga* is optionally
//replaced by *ṣ*.
$if\ X \in \{dviḥ, triḥ, catuḥ\} \wedge y_1(41)$
{
Add *X*|Y to the word_list;
$x_1(23) = 1;$
}

//50. *ataḥ kṛkamikaṁsakumbhapātrakuśā*
//karṇīṣvanavyayasya || 8.3.46 ||
//If *visarga* is preceded by *a* and followed by a form of *kṛ*
//or *kam* or by the words *kaṁsa, kumbha, pātra, kuśā* or
//*karṇī*, then the *visarga* is replaced by *s*.
$if\ x_1(44) \wedge u_1(4) \wedge$ *Y begins with* {*kṛ, kar, kur, kam, kām, kaṁsa, kumbha, pātra, kuśā, karṇī*}
{
$x_1(24) = 1;$
}

//51. *adhaḥ śirasī pade* || 8.3.47 ||
//If the word *adhaḥ* or *śiraḥ* is followed by the word '*pada*',
//then the *visarga* is replaced by *s* optionally
$if\ X \in \{adhaḥ, śiraḥ\} \wedge$ *Y begins with* {*pad*}
{
Add *X*|*Y* to *word_list*;
$x_1(24) = 1;$
}

//Rules 41 to 51 form exceptions to the following rule, Rule
//52, and hence have been handled before it.

//52. *visarjanīyasya saḥ* || 8.3.34 ||
//If *visarga* is followed by a hard consonant, then *s* replaces
//the *visarga*.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

if Flag_kupvoh_sutra_fired = false //8.3.37 is not fired
{
$if\ x_1(44) \wedge (y_1(36) \vee y_1(37) \vee y_1(23) \vee y_1(24) \vee y_1(25))$
{
$x_1(24) = 1;$
}
}

//53. *raṣābhyāṁ no ṇaḥ samānapade* || 8.4.1 ||
// *aṭkupvāṅnumvyavāye'pi* || 8.4.2 ||
// *padāntasya* || 8.4.37 ||
//If *n* is preceded by *ṛ, ṝ, r* or *ṣ* within the same word, and a
//palatal, cerebral, dental, *l, ś* or *s* does not lie between the
//two, and *n* is not the last letter of the word, then *n* is
//replaced by *ṇ*.
for (each letter y in X where y is not the last letter)
{
$if\ (y_1(27) \wedge y_2(2))$ //y is '*n*'
{
$if\ \exists$ a letter x in X preceding y where $((x_1(12) \wedge \neg x_2(2)) \vee x_1(21) \vee x_1(23))$ //x is *ṛ, ṝ, r* or *ṣ*
{
$if\ \nexists$ any letter q between x and y where $q_1(32) \vee q_1(33) \vee q_1(34) \vee (q_1(20) \wedge q_2(3)) \vee q_1(24) \vee q_1(25)$
{
$y_2(1) = 1;$
}
}
}
}
//Repeat the above for Y

//54. *stoḥ ścunāḥ ścuḥ* || 8.4.40 ||
// *śāt* || 8.4.44 ||
//Common name: *ścutva sandhi*
//If a palatal other than *ś* is followed by a dental, or a dental
//is followed by a palatal, then the dental is replaced by the
//corresponding palatal.
$if\ (x_1(32) \wedge \neg x_2(5)) \wedge (y_1(34))$
{
$y_1(32) = 1;$
}
$elseif\ x_1(34) \wedge y_1(32)$
{
$x_1(32) = 1;$
}

//55. *ṣṭunāḥ ṣṭuḥ* || 8.4.41 ||
// *na padāntāṭṭoranāṁ* || 8.4.42 ||
// *toḥ ṣi* || 8.4.43 ||
//Common name: *ṣṭutva sandhi*
//If a dental is followed by a cerebral except *ṣ*, or if the
//specific cerebral *ṭ* is followed by *nām, navati* or *nagarī*, or
//if a cerebral is followed by a dental, then the dental is
//replaced by the corresponding cerebral.
$if\ (x_1(34) \wedge \neg x_2(5)) \wedge (y_1(33))$
{
$x_1(33) = 1;$
}
$elseif\ (x_1(33) \wedge x_2(0)) \wedge Y \in \{nām, navat, nagar\}$
{
$y_1(33) = 1;$
}
$elseif\ x_1(33) \wedge y_1(34)$
{
$y_1(33) = 1;$
}

//56. *yaro'nunāsike'nunāsiko vā* || 8.4.45 ||
//Common name: *anunāsikā sandhi*
//If a consonant other than '*h*' is followed by a nasal, then
//the consonant is optionally replaced by the corresponding
//nasal. The rule is obligatory if the second word is
//'*mayam*' or '*mātram*'.
$if(x_1(36) \vee x_1(37) \vee x_1(38) \vee x_1(39))$
{
$if\ y_1(27)$
{
Add X|Y to the word_list;
$x_1(27) = 1;$
}
$elseif\ Y \in \{maya, mātra\}$
{
$x_1(27) = 1;$
}
}

//57. *aco rahābhyāṁ dve* || 8.4.46 ||
//If *r* or *h* is followed by any consonant other than *h* and
//preceded by a vowel, then the consonant is duplicated
//within a word.
for (each set of consecutive letters *u,x,y* in *X*)
{
$if\ (x_1(21) \vee x_1(26)) \wedge (y_1(2) \wedge \neg y_1(26)) \wedge u_1(0)$
{
$z = y;$
Add z after x;
}
}
for (each set of consecutive letters *u,x,y* in *Y*)
{
$if\ (x_1(21) \vee x_1(26)) \wedge (y_1(2) \wedge \neg y_1(26)) \wedge u_1(0)$
{
$z = y;$
Add z after x;
}
}

//58. *anaci ca* || 8.4.47 ||
// *dīrghādācāryāṇām* || 8.4.52 ||
//If any consonant other than *h* is preceded by a short vowel

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

```
//and followed by anything other than a vowel, then the
//consonant is doubled within a word.
for (each set of consecutive letters x,y,w in X)
{
    if x1(8) ∧ (y1(2) ∧ ¬y1(26)) ∧ (w1(1) ∨ w1(2) ∨
    w1(3))
    {
        z = y;
        Add z after x;
    }
}
for (each set of consecutive letters x,y,w in Y)
{
    if x1(8) ∧ (y1(2) ∧ ¬y1(26)) ∧ (w1(1) ∨ w1(2) ∨
    w1(3))
    {
        z = y;
        Add z after x;
    }
}
```

//59. *jhalām jaś jhaśi* || 8.4.53 ||
//Common name: *jaśtva sandhi*

```
//If a non-nasal mute, sibilant or aspirate is followed by
//Column3 or Column4, then the first letter is replaced by
//the corresponding Column3 letter.
for (each set of consecutive letters x, y in X)
{
    if (y1(38) ∨ y1(39))
    {
        if x1(36) ∨ x1(37) ∨ x1(38) ∨ x1(39) ∨ x1(26)
        {
            x1(38) = 1;
        }
        elseif x1(23) ∨ x1(24) ∨ x1(25)
        {
            x1(38) = 1;
            if x1(23)
            {
                x2(1) = 1;
            }
            elseif x1(24)
            {
                x2(2) = 1;
            }
            elseif x1(25)
            {
                x2(3) = 1;
            }
        }
    }
}
//Repeat the above for Y
```

//60. *khari ca* || 8.4.55 ||
//Common name: *cartva sandhi*

```
//If a non-nasal mute or sibilant is followed by a
//hard consonant, then the first letter is replaced by the
//corresponding Column1 letter.
if (x1(36) ∨ x1(37) ∨ x1(38) ∨ x1(39) ∨ x1(23) ∨
x1(24) ∨ x1(25)) ∧ (y1(36) ∨ y1(37) ∨ y1(23) ∨
y1(24) ∨ y1(25))
{
    x1(36) = 1;
}

//Check internally in each word too
for (each set of consecutive letters x, y in X)
{
    if y1(36) ∨ y1(37) ∨ y1(23) ∨ y1(24) ∨ y1(25)
    {
        if (x1(36) ∨ x1(37) ∨ x1(38) ∨ x1(39))
        {
            x1(36) = 1;
        }
        elseif x1(23) ∨ x1(24) ∨ x1(25)
        {
            x1(36) = 1;
            if x1(23)
            {
                x2(1) = 1;
            }
            elseif x1(24)
            {
                x2(2) = 1;
            }
            elseif x1(25)
            {
                x2(3) = 1;
            }
        }
    }
}
//Repeat the above for Y
```

//61. *anusvārasya yayi parasavarṇaḥ* ||8.4.58||
//Common name: *parasavarṇa sandhi*

```
//If anusvāra is followed by a semi-vowel or a mute, then
//the anusvāra is replaced by the nasal equivalent of the
//second letter.
if x1(43)
{
    if y1(20)
    {
        x1(22) = 1;
        x2 = y2;
    }
    elseif y1(36) ∨ y1(37) ∨ y1(38) ∨ y1(39)
    {
        x1(27) = 1;
        x2 = y2;
    }
}
```


http://sites.google.com/site/ijcsis/
ISSN 1947-5500

```
//62. torli || 8.4.60 ||
//Common name: parasavarṇa sandhi
//i) If n is followed by l, then n is replaced by nasal l.
//ii) If a dental other than n and s is followed by l, then
//the dental is replaced by l.
if (x1(27) ∧ x2(2)) ∧ (y1(20) ∧ y2(3))
{
    x1(22) = 1;
    x2 = y2;
}
elseif (x1(34) ∧ ¬(x2(4) ∨ x2(5))) ∧ (y1(20) ∧ y2(3))
{
    x = y;
}

//63. jhayoho'nyatarasyām || 8.4.62 ||
//Common name: pūrvasavarṇa sandhi
//If a non-nasal mute is followed by h, then h is optionally
//replaced by the Column4 letter corresponding to the non-
//nasal mute.
if (x1(36) ∨ x1(37) ∨ x1(38) ∨ x1(39)) ∧ y1(26)
{
    Add X|Y to the word_list;
    y1(39) = 1;
    y2 = x2;
}

//64. śaścho'ṭi || 8.4.63 ||
//Common name: chatva sandhi
//If a non-nasal mute is followed by ś which is in turn
//followed by a vowel, semi-vowel or nasal, then ś is
//optionally replaced by ch.
if(x1(36) ∨ x1(37) ∨ x1(38) ∨ x1(39)) ∧ y1(25) ∧
(w1(0) ∨ w1(1) ∨ w1(27))
{
    Add X|Y to the word_list;
    y1(40) = 1;
}
//65. halo yamāṁ yami lopaḥ || 8.4.64 ||
//If a semi-vowel or nasal is preceded by a consonant and
//followed by the same semi-vowel or nasal letter, then one
//of the duplicate letters is dropped.
if u1(2)
{
    if (x1(20) ∨ x1(27)) ∧ y == x
    {
        delete x;
    }
}

//66. jharo jhari savarṇe || 8.4.65 ||
//If a non-nasal mute or sibilant is preceded by a consonant
//or semi-vowel and followed by a homogeneous mute or
//sibilant, then one of the duplicate letters is optionally
//dropped, within a word
for (each set of consecutive letters u,x,y in X)
{
    if u1(1) ∨ u1(2)
    {
        if (x1(36) ∨ x1(37) ∨ x1(38) ∨ x1(39) ∨
        x1(23) ∨ x1(24) ∨ x1(25)) ∧ (x1 == y1)
        {
            Add X|Y to the word_list;
            delete x;
        }
    }
}
// Repeat the above for Y
```

C. The Search Engine

Algorithm SandhiProcessor may be used to generate alternative forms of a given search word. The following algorithm is used to generate all possible alternative forms of a given word, by providing possible word forms before and after the word so that *sandhi* rules get triggered. All these word forms are searched for in the E-text.

```
Algorithm GenerateAllWordForms (Z)
{
//Z is the search word.
//{WordForms} denotes the set of word forms generated by
//the algorithm and is initially the null set.
    Add Z to {WordForms};
    X = Z;
    for (each y in {vowels, semi-vowels, consonants})
    {
        Add SandhiProcessor(X, Y) to {WordForms};
    }

    for (each Y ∈ {WordForms})
    {
        for (each X in {vowels, semi-vowels, consonants, ḥ, ṁ,
        #})
        {
            Add SandhiProcessor(X, Y) to {WordForms};
        }
    }
}
```

VI. Conclusion

The schema developed in this work presents a simple yet unique and efficient method to process the *sandhi* aphorisms of Pāṇini. The letter representation scheme is binary, and hence all the checks are implemented as bit-level operations and simple bit-set and bit-unset operations suffice to carry out the *sandhi* transformation. The efficiency is further enhanced by the division of a letter representation into two parts and the consequent reduction of the transformation process to a shifting of category. Further, this pattern of solving the sandhi construction problem is unprecedented in the literature. Thus, representation schema and the results of the *sandhi*-processing algorithm represent an efficient computational model to process Sanskrit euphonic conjunctions. It must be mentioned here that some rules such as those with regard to *prakṛtibhāva sandhi*

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

(non-transformational *sandhi*) have not been presented above. However, it is clear that their implementation is only an extension of the algorithm presented in this work that does not require any new schema.

The representational schema has been reduced further from [4] in this work, since many more *sandhi* rules have been incorporated here. The optimality of the schema is clear from the simplicity of the rule representation.

The final algorithm presented in this work, which uses this *sandhi* processor for word searches in E-texts is the first of its kind in the literature with regard to Sanskrit.

The use of the *sandhi* processor for searching ensures comprehensiveness of the search, while the efficiency of the *sandhi* processing method presented in this work ensures that search speeds are not compromised due to the increase in the number of words to be searched for. This was confirmed in the implementation of the algorithm.

The algorithms presented in this work have been tested for use with Sanskrit E-text in *Devanāgarī* script after a conversion engine converted *Devanāgarī* Unicode to Latin Unicode E-text.

REFERENCES

[1] Dīkṣita Bhaṭṭoji, ***Siddhānta-kaumudī***, Translated by Śrīśa Candra Vasu, Volume 1, Motilal Banarsidas Publishers, Delhi, 1962.
[2] Vāmana & Jayāditya, ***Kāśikā***, with the subcommentaries of Jinendrabuddhi, Haradatta Miśra and Dr. Jaya Shankar Lal Tripathi, Tara Book Agency, Varanasi, 1984.
[3] Rama N., Meenakshi Lakshmanan, *A New Computational Schema for Euphonic Conjunctions in Sanskrit Processing*, International Journal of Computer Science Issues, Vol. 5, 2009, pp 43-51 (ISSN print: 1694-0814, ISSN online: 1694-0784).
[4] Kasmir Raja S. V., Rajitha V., Meenakshi Lakshmanan, *A Binary Schema and Computational Algorithms to Process Vowel-based Euphonic Conjunctions for Word Searches*, International Journal of Applied Engineering Research, Volume 9,Number 20(2014) pp. 7127-7141 (ISSN print: 0973-4562, ISSN online: 1087-1090).

Websites

[5] Göttingen Register of Electronic Texts in Indian Languages (GRETIL), *gretil.sub.uni-goettingen.de/gretil.htm*
[6] Sanskrit Documents, sanskritdocuments.org
[7] Indology: Resources for Indological Scholarship, *indology.info/etexts/archive/ etext*
[8] TITUS, titus.uni-frankfurt.de/indexe.htm
[9] Muktabodha Indological Text Collection & Search Engine, *muktalib5.org/digital_library_secure_entry.htm*

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Classification of Sleep Stages Using Neural Network Based on EEG and EOG signals

[1]Shreya Garg
Jaypee Institute of Information Technology
Noida, India

[2]Vijay Kahre
Jaypee Institute of Information Technology
Noida, India

Abstract— This paper introduces an algorithm for different sleep stages classification. The algorithm consists of wavelet packet transformation (WPT) which is applied to 30 seconds long epochs of EEG and EOG recordings to provide time-frequency information, a feature generator to quantify the information and reduce the data set size, and then artificial neural networks for doing optimal classification. This led to a classification method with efficiency of 90.41 percent.

Keywords-component; Neural Network (NN), Electroencephalograph (EEG), Electrooculorgraph (EOG), Polysomnography (PSG), Wavelet Packet Transform(WPT)

I. INTRODUCTION

Polysomnography (PSG) is a medical diagnostic test used to monitoring the activity of various organ systems during sleep. It is used to record various signals Electroencephalogrm(EEG), Electrooculorgrah(EOG), Electromayogram (EMG), Electrocardiogram(ECG), Respiratory Airflow, Respiratory Effort Indicators and Pulse Oximetry. PSG used for the diagnosis of apnoea-hypopnoea and upper airway resistance syndromes and a variety of other sleep conditions related to daytime sleepiness that cannot be classified as breathing disorders, such as restless leg syndrome, disorders during REM sleep and other parasomnias [1-2].

II. MATERIAL AND METHODOLOGY

A. *Data Acquisition*

In this study data was acquired using eights Ag- Agcl electrodes from sleep recording of 7 hours 27 minutes. Six electrodes were used to record electroencephalogram (EEG). Electrodes are placed according to international standard 10-20 system at frontal, central and occipital lobes of the brain as shown in Figure1[3]. Two electrodes were used for Electro-oculorgram (EOG) as shown in Figure2. One electrode is placed one cm above the outer canthus of the right eye and other electrode is placed one cm below the outer canthus of the left. These electrodes pick up the activity of the eyes in virtue of the electro potential difference between the cornea and the retina [2]. From sleep recording obtained data set as shown in Figure3.

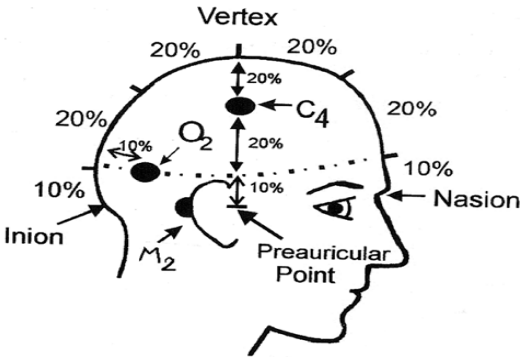


Fig 1 .EEG electrode placement according to 10-20 system

Fig 2: EOG Electrode placement

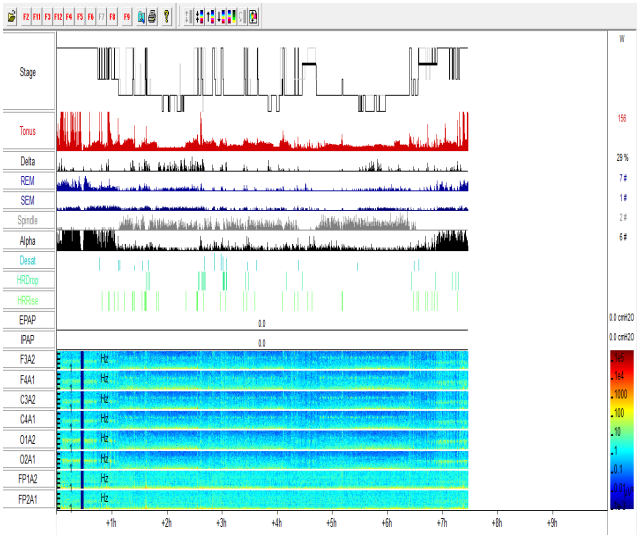


Fig 3: Summary of a sleep study

1.380 wake,
2.120 N1,
3.1006 N2,

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

4.144 N3,
5 138 Rem stages,
Data was also obtained from electrodes EOG left and EOG right with 894 data sets from each electrode.

B. Wavelet Packet Transform

The wavelet Packet transform is a method using which given signal or a waveform can decompose into its frequency and location elements. This is done by using a mother wavelet defined by $\psi(x)$. At each stage, Details as well as Approximations are further decomposed into low and high frequency signal components and make wavelet packet decomposition tree[4-5]. This leads to an equal number of the approximation and detail coefficients. For our analysis, mother wavelets db 20 and db 2 from daubechies family were used for EEG and EOG signal (it resembles the EEG and EOG amplitudes most).

C. Sleep stages and structure

In humans, sleep stages have a specific pattern of frequency content. The EEG is divided in 5 frequency bands which are,
Delta 0 – 4 Hz
Theta 4 -8 Hz
Alpha 8 – 13 Hz
Beta1 13 – 22 Hz
Beta2 22 – 35 Hz

The following sleep stages are defined:
Stage awake: Signal with continuity alpha activity.
Stage 1: alpha activity is disappear and presence of low beta and theta activity.
Stage 2: Presence of delta activity (less than 20 %) and K-complexes and spindles. K-complex waves have low frequency 1.0 Hz and of at least 75 mV. Spindles are waves in the range 11-15 Hz with time duration of more than 0.5 seconds. There is no criterion about the amplitude of a spindle.
Stage 3: Delta activity lies in the range of (20 % to 50 %)
Stage REM: It's low amplitude with little theta waves. REM and awake signals have a similar shape, but REM has little alpha activity.

D. Feature Extraction

For feature extraction wavelet packet transform was used for EEG and EOG signals. By applying wavelet packet transform (WPT) on raw EEG signal (sampled at 100Hz) decomposed up to 6 levels. Wavelet coefficients were obtained in following 7 frequency bands.

1. 0.4 - 1.55 Hz ,K-complexes + Delta
2. 1.55 - 3.2 Hz, Delta
3. 3.2 - 8.6 Hz, Theta
4. 8.6 - 11.0 Hz, Alpha
5. 11.0 - 15.6 Hz, Spindle
6. 15.6 - 22.0 Hz, Beta1
7. 22.0 - 37.5 Hz, Beta2

Small-frequency waves (delta wave, K- complexes) have broader time resolution and high-frequency waves (spindle, alpha waves) have finer time resolution.
For every 30 seconds epoch has taken from the central EEG electrode C3 and C4 and calculates the mean quadratic value of the WP coefficients for each of the 7 bands. features for the epoch. Additionally we defined 6 more features based on total energy and the ratio of different energy values:

8. Total Energy of the 7 bands

$$E = \sum_{i=1}^{7} E_i$$

9. Ratio (E1 + E2)/E8 Percent Delta Activity
10. Ratio E4/E8 Percent Alpha Activity
11. Ratio (E1+E5)/E8 Percent KK and Spindle
12. Ratio E4/E3 Ratio Alpha/Theta
13. Ratio (E1+ E2)/E3 Ratio Delta/Theta ,

Wavelet packet transform was used to decompose EOG signal sampled at 100Hz up to 3 levels and obtained wavelet coefficients lie in frequency band 0-12.5 Hz .Hence define 2 more features as follows,

14. E14 EOG left energy
15. E15 EOG right energy.

E. Classification

The main advantage of choosing artificial neural network for classification was due to fact that ANN's could be used to solve problems, where description for the data is not computable. ANN could be trained using data to discriminate the feature [6-8]. For classification a two layer neural networks was used for the instance a topology of {25, 1} indicate a 15 input, 25 neurons in hidden layer and one output layer. After feature generation, data points are fed into a cascaded system of fully connected feed forward back propagation neural networks with momentum in weight as shown in Figure4:

1. **Network 1(NET1):** This is to classify whether signal is sleep stage or wake stage. If the signal is classified as sleep stage, data is fed into network 2.
2 **Network 2(NET2):** This is to classify whether the signal is REM or N- REM. If the signal is classified as N-REM stage, data is fed into network
3. **Network 3(NET3**): This is to classify whether signal is N1 or not. If the signal is not N1 stage, data is fed into network 4.
4. **Network 4(NET4):** It distinguishes whether a signal is N2 or N3.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

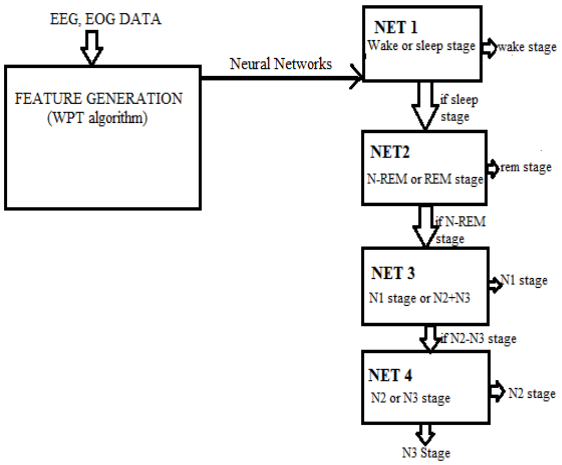


Fig 4: sleep stages Classification using neural network

III. RESULTS:

In this study for classification of sleep stages, wavelet packet transform were used for feature extraction of EEG signal and EOG signal. Mother wavelet db20 and db 2 (Daubechies Family) was used for decomposition of respectively EEG and EOG signal. For classification, neural network with Gradient descent Back propagation with momentum were used. Percentage of accuracy for C3 and C4 electrodes of each network is shown in the Table 1a and Table 1b.

Table 1a: Results for data sets obtained from electrode C3

Network	Number of testing data sets	Number of sets correctly identified	Percentage accuracy
Network 1	314	284	90.44
Network 2	224	188	83.92
Network 3	205	191	93.17
Network 4	150	110	73.33

Table 1b: Results for data sets obtained from electrode C4

Network	Number of testing data sets	Number of sets correctly identified	Percentage accuracy
Network 1	314	299	95.22
Network 2	224	198	88.39
Network 3	205	186	90.73
Network 4	150	131	87.33

IV. CONCLUSION:

A method for doing automatic sleep stage classification has been proposed and tested on a data set obtained from Neurology and Sleep Centre, New Delhi. As previously mentioned, wavelet packet transform was used to perform adaptive time frequency resolution. Neural network with Gradient Descent back propagation with momentum method was used and average accuracy 90.41 % obtained. Obtained results are very encouraging and show this method could be used to classify the sleep stages. As a future work, we plan to do Performance optimization for feature extraction and classification and then incorporate this work into real time monitoring system that acquired and analyzes the EEG and EOG signal of the subject during sleep.

Acknowledgements: The authors would like to acknowledge their gratitude to the staff of neurology and sleep centre, New Delhi for help in carrying out the experiments.

V. REFRENCES

[1] A. Minaritzoglou, E. Vagiakis, Sleep Polysomnography: Recent data on Procedure and Analysis, PNEUMON Number 4, Vol. 21, 2008
[2] A. Rechtschaffen, and A.Kales, A Manual of Standardized Terminology, Technique and Scoring System for Sleep Stages of Human Subjects, Public Health Service, U.S, Government Printing Office, Washington, DC, 1968
[3] R. W. Homan, J. Herman, and P. Purdy,"Cerebral location of international 10–20 system electrode placement Localization cerebrale des electrodes placees selon le system international 10–20,Electroencephalography and Clinical Neurophysiology,66(4):376,382 ,1987
[4] S. Mallat, A theory for multiresolution. signal decomposition: The wavelet representation, IEEE Trans. Patt. Recog. And Mach. Intell., 11(7) : 674-693,1989
[5] Marc Jobert, Christian Timer, Eric Oiseau and Harmut Schulz, Wavelets - a new tool in sleep biosignal analysis, Journal of Sleep Research (3): 223-232,1994
[6] E. Oropesa, H. L. Cycon and M. Jobert, Sleep Stage Classification using Wavelet Transform and Neural Network, International Computer Science Institute,TR-99-008, March 1999
[7] M.Malini and K.Subba Rao, Analysis of EOG Signal using Haar Wavelet, International Symposium on Devices MEMS, Intelligent Systems & Communication (ISDMISC) (2):28-31,2011.
[8] M. Moreira and E. Fiesler, Neural Networks With Adaptive learning rates and momentum terms, IDIAP Technical report, October 1995.

AUTHORS PROFILE

Vijay Khare: He did his PhD in Bio Signal Processing at the Indian Institute of Technology, Delhi. He did his M.Tech in Instrumentation & Control, from NSIT Delhi. He is currently, with the Dept. Electronics and Communications Engineering at the Jaypee Institute of Information Technology. His research interests are Neural Networks, Brain Computer Interfacing, and Biomedical Signal Processing.

Shreya Garg: She did her B.Tech from Jaypee Institute Of Information technology. Her research areas are neural network and biomedical signal.

An Ultra Low Power and High Throughput FPGA Implementation of SHA-1 Hash Algorithm

Shahzad Khan
Department of Computer Science
Shaheed Benazir Bhutto University(SBBU)
Khyber PukhtunKhwa, Pakistan

Zain-ul-Abideen
Department of Communications System Engineering
School of Electrical Engineering and Computer Sciences(SEECS), NUST, Islamabad, Pakistan

Shahzad Shahid Paracha
Department of Communication System Engineering
School of Electrical Engineering and Computer Science(SEECS), NUST, Islamabad, Pakistan

Abstract *- In this paper, we present a low power and highly parallel SHA-1 architecture which is considered as extremely iterative in nature specifically suitable for power sensitive applications. That is achieved by first identifying non dependent operations among the consecutive iterations of the algorithm and then aligning them for their execution in a highly parallel way. Consequently, when iteration completes, some other iterations also get completed and only a few of their dependent operations are left. By using this approach we were able to perform up to four SHA-1 iterations simultaneously resulting an increase in its throughput approximately by four times. We also explain how our results critically effect in lowering down the power consumption of the design.*

Keywords - Cryptography, Hash function, FPGA implementation

I. *INTRODUCTION*

Secure Hash Algorithm (SHA) is used to provide message authentication in widely used security applications and protocols. The fast growing technology and the evolution of the protocols resulted in high-performance applications. In-order to be compatible with the portable, small sized and less power consuming devices the throughput of SHA needs to be increased. SHA-1 is mostly used by the security applications such as IPSec and WTLS.

In order to provide both source authentication and data integrity HMAC (Hashed- Message Authentication Code) is used, which employs either iterative hash functions MD5 or SHA-1. The iterative hash function takes variable length input and gives out a fixed length output called as Message Digest after many iterations to introduce more confusion and diffusion. HMAC is used as a tool to provide source authentication and message integrity in both wired and wireless networks. There are various standards that specify the security need in wired networks and mobile services such as IPSec for Local Area Network, WTLS for WAP and 802.16 for Metropolitan Area Network. These newly evolved communication standards all require an optimized and efficient HMAC implementation in order to achieve data integrity and source authentication. The optimized HMAC can be achieved if we make the iterative hash function work efficiently.

The attempts so far have been in decreasing the implementation size of the SHA-1 algorithm through re-use of the components and reconfiguration [4]. There have been further techniques in increasing the throughput of the hash functions employing the concepts of parallelism and pipelining. One of the proposed designs that aimed to achieve both higher throughput and smaller area coverage is presented in [1] and [2] employing pipelining and re-use techniques. Another improved design focused more on increasing throughout by 37% without significant area penalty [3].

Among the family of hash algorithms SHA-1 and MD5 both have some key advantages over one another such as MD5 is faster in implementation whereas SHA-1 is more secure. Our main focus shall be on SHA-1 for its comparatively stronger security and common use in security standards and protocols. The recent successful attacks on MD5 algorithm [8] and smaller key size of 128 bits makes it more vulnerable and less favorable to be focused for optimization.

This paper carries further the process of increasing throughput of SHA-1 hash algorithm through the deeper analysis of the optimized design presented in [3] and improved in [4]. Design is also analyzed for performance and size considerations. The proposed design increases the throughput by 15% as compared to the last optimized design presented in [4] alongwith the low power considerations of the implementation. Our approach is to

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

analyze and test the designs presented in [3] referred by us as SHA1in1 and in [4] referred to as SHA2in1 alongwith our proposed design referred to as SHA4in1 on Xilinx simulator for reliable figures verifying our claim.

This paper is organized in following way. Section 2 describes the basic SHA algorithm design and its constituent components and functions. Section 3 presents the summary of some previous implementations of SHA-1 with different performance features. Our proposed design with throughput and area consideration is presented in Section 4. Section 5 presents performance results and Section 6 presents some key comparisons with other implementations in terms of throughput, power and area size. Finally the last section 7 provides conclusion.

II. *OVERVIEW OF SECURE HASH ALGORITHM (SHA)*

The standard architecture of SHA-1 Hash function is described in Secure Hash Standard [7] and basic functionality of SHA-1 is described in which 4 rounds, each of 20 iterations are used to operate on 512 bit block of data to generate fixed length Message Digest of 160 bits, hence 80 iterations in total are performed in order to get the required output. Here, a round represents a stage. Iterations within same round have exactly identical functionality but they differ with different rounds due distinct non-linear functions F_i used within each round.

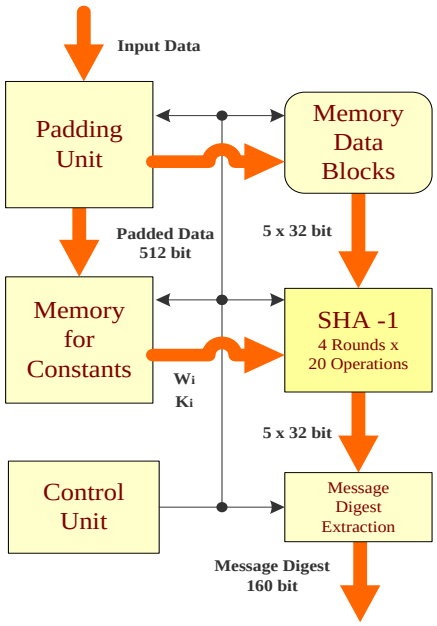


Fig 1: *SHA-1 Block Diagram.*

The design is fed with five 32bit variables a, b, c, d and e, the constants K_t and W_t, of 32 bits each. The variables a, b, c, d and e act as Initial Vector to SHA-1 hash function.

The constant K_t is a fixed value and is distinct for each round. The constant W_t is derived from the original messages and is different for every iteration of one round. Each round runs for 20 iterations with one non-linear function, after 20^{th} iteration the output is fed to the 2^{nd} round for further 20 iterations with different non-linear function. The output is then fed to round 3 and 4 both having distinct nonlinear functions for 20 iterations each. In this way the value is processed by 4 rounds each performing 20 iterations with different nonlinear functions F_i. In total there are 80 iterations performed in 4 rounds to produce the Message Digest of 160bit.

In each round data process is based on non-linear function F_i and inputs b, c and d [7]. the result of this all process is added to the fourth input with the constant K_t and the messages block W_t. Some shifting and rotation operations are also performed during all this process. This process is repeated in each iteration when the resultant output of round one is fed to the 2^{nd}, 3^{rd} and 4^{th} round, respectively, finally resulting the Message digest of 160 bits.

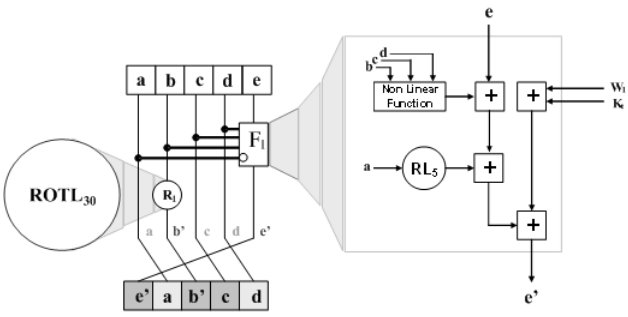


Fig 2: *Internal design on SHA-1 core*

We need some memory storage in order to store all these constants and variables. MS Ram is used to store message schedule W_t of t^{th} 32-bit word of the padded message for t iterations. The constant array of K and the values of the initial vectors (IV) from *a* to *e* are also stored. The large number of operations kept the throughput of SHA-1 hash function very low. Many Techniques have been presented that describe improved SHA-1 implementation. Some of them are as follows.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

III. *RELATED WORK*

There have been many efforts on improving the performance of SHA-1 in terms of speed, throughput, power consumption and area size. The ultra high speed architecture of SHA-1 proposed in [6] raises the throughput to 1,4 Gbps with relatively better performance and superior performance/area ratio but the room for further improvement remains. The reconfigurable architecture in [1] employs pipelined design improving throughput upto 1,7 Gbps for SHA1, and 2,1 Gbps for MD5. The achieved performance is greater than that of MD5. [4] improves the throughput further by 55% while keeping power dissipation lower. Further the high throughput implementation of SHA-1 in [4] raised the limit of throughout to maximum 2,8 Gbps exceeding the limit by 37% without significant area penalty. This [4] is the highest throughput achieved as yet. The concept of parallelism is the significant key feature of improvement that is exploited in favour of throughput in [5].

IV. *OUR PROPOSED DESIGN*

Secure Hash Standard [7] describes the functionality and architecture of the SHA-1 hash algorithm as shown in Fig 1. The SHA-1 core comprises of 4 rounds having 20 iterations each, with 80 rounds in total that gives a Message Digest as output. Each Message Digest block comprises of 160 bits as per the standard [7] and the message block is taken as input divided in the groups of 512 bits for each of the 4 rounds. Each chunk of 512 bits is further broken down into 16 words of 32 bits each. For total 80 iterations in 4 rounds, total of 80 words are given as input along with a key. The 32 bit words from 17 upto 80 are derived from first 16 words taken from 512 bit block.

From the previous implementations of SHA-1 we adapt the following equation and working further on this equation we are able to derive an optimized solution.

$$e_f = a' = ROTL_{30}(a)$$
$$d_f = e'' = ROTL_{30}(ROTL_5(a) + f_1(b, c, d) + e + W_1 + K_t)$$
$$c_f = d'' = ROTL_{30}(ROTL_5(ROTL_5(a) + f_1(b, c, d) + e + W_1 + K_t) + f_2(a, ROTL5(b), c) + d + W_2 + K_t)$$

$$b_f = c' = ROTL_5(ROTL_5(ROTL_5(a) + f_1(b, c, d) + e + W_1 + K_t) + f_2(a, ROTL5(b), c) + d + W_2 + K_t) + f_3((ROTL_5(a) + f_1(b, c, d) + e + W_1 + K_t), ROTL_{30}(a), ROTL_{30}(b)) + c + W_3 + K_t$$

$$a_f = b'' = f_4((ROTL_5(ROTL_5(a) + f_1(b, c, d) + e + W_1 + K_t) + f_2(a, ROTL_5(b), c) + d + W_2 + K_t), ROTL_{30}(ROTL_5(a) + f_1(b, c, d) + e + W_1 + K_t), ROTL_{30}(a)) ROTL_5(ROTL_5(ROTL_5(ROTL_5(a) + f_1(b, c, d) + e + W_1 + K_t) + f_2(a, ROTL5(b), c) + d + W_2 + K_t) + f_3((ROTL_5(a) + f_1(b, c, d) + e + W_1 + K_t), ROTL_{30}(a), ROTL_{30}(b)) + c + W_3 + K_t) + ROTL_{30}(b) + W_4 + K_t$$

where $f_1 = f_2 = f_3 = f_4$ **(1)**

The proposed design, as shown in Fig 2, works similar to [4] but with an enhancement of parallelism of four sequential iterations within each round instead of two iterations, which means that the values of the variable a_f, b_f, c_f, d_f and e_f each can be derived from the output produced by the round function as shown in equation (1).

For the first two iterations we wait for the output of the non-linear functions and as we get the outputs of both the initial two iterations we then precede to 3rd and 4th iterations in parallel process, the values are assigned to them from first two iterations resulting in output of 4 iterations. We divide our solution in 6 delays and 4 parallel iterations as shown in Fig 4. During each delay we get the intermediate outputs and after 6 delays from D_1 to D_6 we get the complete result for the four iterations. This result is looped back to the 1st iteration block as an input and similar operation is performed and output is generated repeatedly.

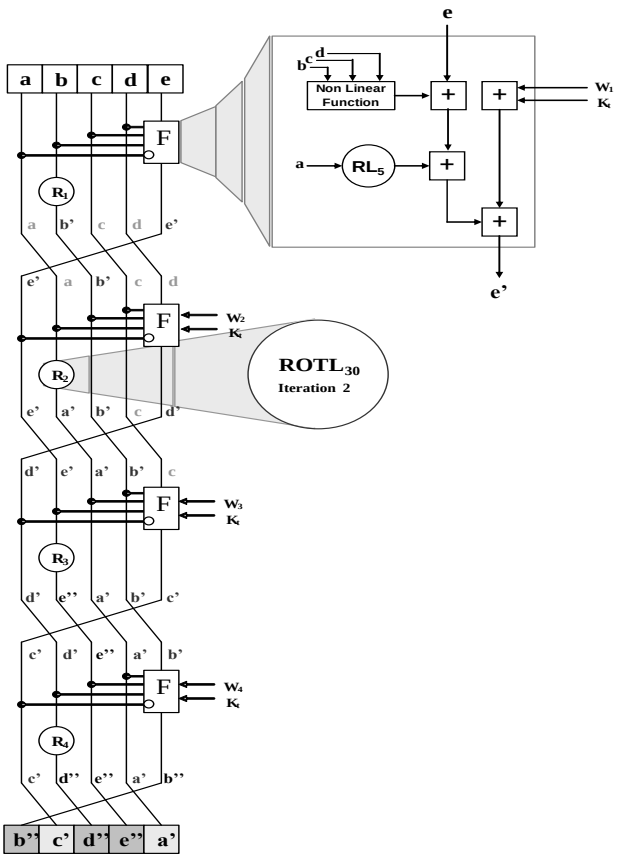


Fig 3: *Four consecutive SHA-1 blocks of the same round*

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

We run this parallel block of 4 iterations five times in order to get the result of complete 20 iterations in the first design. In this technique with the help of parallelism we derive the results in lesser time as compare to the early implementations. Therefore we conclude that with the insertion of few additional levels we make our throughput increased to a greater extent as compare to early SHA-1 implementations.

The throughput of the proposed design would simply be more than 2,8 Gbps as we have introduced parallelism into the design resulting in reduction of four iterations into just one by adding more computational logic. The expression to calculate the throughput is given in equation (2) :

$$Throughput = \#\ bits\ .\ f_{op} / \#\ operations \qquad (2)$$

V. *PERFORMANCE RESULTS AND DISCUSSION*

There are two approaches to achieve optimized and efficient SHA-1 hash function.

1. Minimize SHA-1 implementation size.
2. Alternative design approach.

In first technique the most common solutions are operation rolling loop and reconfiguration where as in the latter there are alternative designs that have been proposed to achieve high throughput and the technique used for it is parallelism. We code the three designs in Verilog HDL as listed on Table 1, namely SHA1in1, SHA2in1 and our proposed design SHA4in1, and simulate them using Xilinx FPGA Device v3200efg1156 to see what values we get for minimum period.

The SHA4in1 SHA2in1 and SHA1in1 design codes have been tested against the standard test values given in [7] FIPS PUB 180-1 with two standard test cases examples.

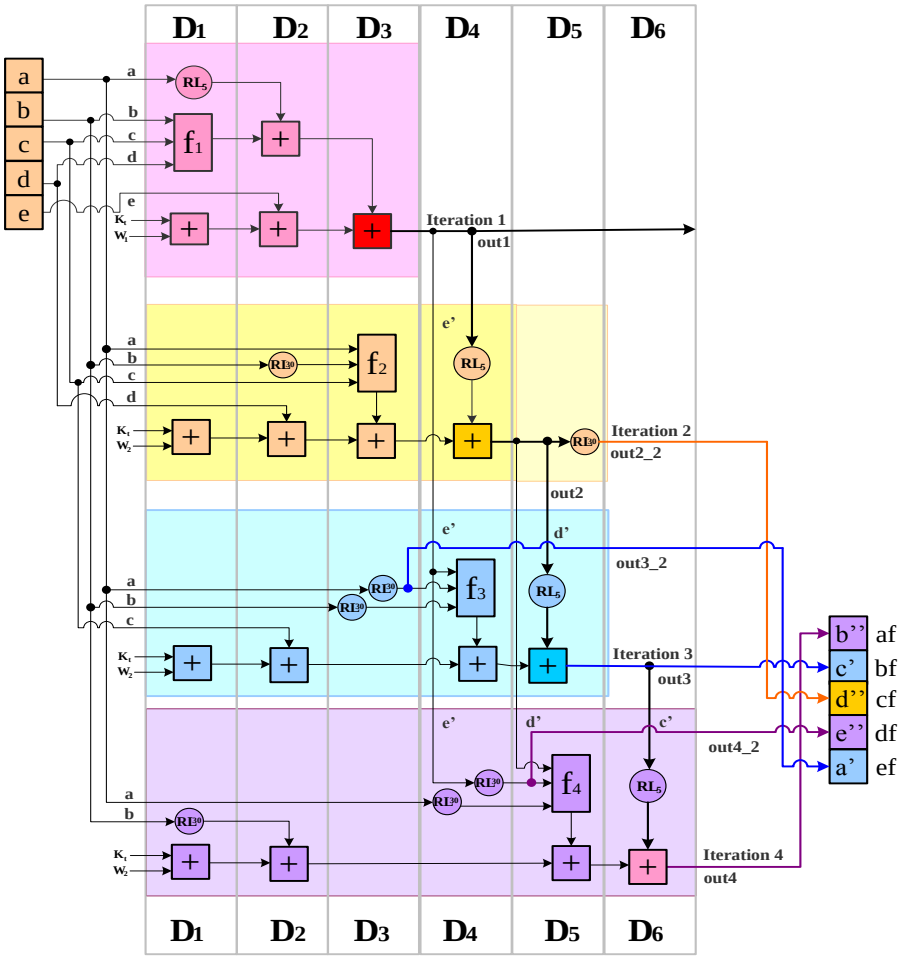


Fig 4: *Proposed Design of parallelized four parallelized iteration blocks in one round*

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The minimum period SHA1in1 design comes out to be 11,62 ns with maximum operating frequency at 86,1 MHz. The value of minimum period increases in case of SHA2in1 being 17,47 ns with reduced operating frequency 57,3 MHz and giving same throughput. Our total number of iterations are 20 and for each iteration minimum time is 29,76 ns that gives 148,8 ns for single round. For that of SHA1in1 it takes 20 iterations, for SHA2in1 it takes 10 and for our proposed design SHA4in1 only 5 iteration give the same result.

Table 1: Comparison of minimum period, maximum operating frequencies and throughputs for each of the design tested

SHA-1 Tested Designs	**Data from Xilinx Simulation of XILINX - v3200efg1156 Device**		
	Minimum Period (ns)	**Operating Frequency (MHz)**	**Time for Single Round (ns)**
Tested SHA1in1	11,62	86,1	20 x 11,62 = 232,4
Tested SHA2in1	17,47	57,3	10 x 17,47 = 174,7
Proposed SHA4in1	**29,76**	**33,6**	**5 x 29,76 = 148,8**

The time for single round for SHA1in1 comes out to be 232,4 ns, for SHA2in1 it is equal to 174,7 ns that is 25% less as compared to that of SHA1in1. The time for single round for SHA4in1 takes 148,8 ns, that is 36% less than that of SHA1in1 and 14.8% less than that of SHA2in1. Therefore our proposed solution SHA4in1 is much faster than the previous designs of SHA1in1 and SHA2in1 giving us optimal solution both in terms of power consumption and speed or throughput.

A. AREA EFFICIENT SHA-1 IMPLEMENTATION

The architecture of the SHA-1 core is area-efficient and requires 4 basic operation blocks each with a different non-linear function. This can be achieved with a counter and temporal register. Each operation block will then be iterated 20 times in order to achieve the desired results. The output of each iteration is stored in temporal register that is fed as an input for the next round making a very simple architecture implementation of SHA-1. This approach is area efficient but the throughput is very low due to the large number of clock cycles needed in producing the Message Digest of 160 bits.

B. REUSING SHA-1 HARDWARE

Each round are similar apart from the non-linear function they use. The technique includes all four non-linear functions to group together in one operation block with the help of a multiplexer the required non-linear function is selected against the time instance *t* for each round. In this implementation the numbers of operation blocks are reduced with addition of a multiplexer and remaining three non-linear functions.

C. Introducing Parallelism

In this technique four parallel stages are introduced to SHA-1 implementation. The basic architecture of SHA-1 requires 4 different non-linear functions for each of the 4 different rounds and each round contains 20 iterations. Each round is assigned one stage and re-use of the operations blocks from the previous designs helps in reduced area requirement. With the help of this procedure complete operations of the four rounds in parallel are achieved.

D. LOW POWER AND HIGH-SPEED IMPLEMENTATION OF SHA-1

This approach identifies the core elements that had impact on throughput and then manipulates those elements for better results. The technique focuses on the special SHA-1 property i.e. the parallelism introduced in the operations of each round of SHA-1. The design is modified in such a way that the consecutive operations of a round are performed in parallel. This involves inclusion of an addition level to the critical path.

This reduces the maximum operating frequency due to increase in minimum period but provides with a higher throughput as compared to early implementations. The factor of power dissipation is also considered and this mechanism provides the solution with lower power dissipation and higher throughput.

We reduce the number of iterations in each round from 20 to 5 and thus the total cycles for four rounds from 80 to 20 which is the significant development. There is 75%

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

reduction in number of cycles as compared to [3] and 50% reduction as compared to [4].

The dynamic power dissipation significantly comes down due to reduction in operating frequency and the no. of cycles of operations in each round.

Table 2: *Operating Frequencies, Throughput and Area Comparisons*

SHA-1	XILINX FPGA - v3200efg1156 Device		
	Operating Frequency (MHz)	**Throughput (Mbps)**	**Area (CLBs/Dffs)**
[1]	71	1731,0	1018 / 1036
[2]	72	1843,2	878 / 1735
[3]	98,7	2527,0	950 / 1164
[4]	55	2816,0	- / -
[5]	42,9	119,0	1004 / -
[6]	55	1339,0	2245 / -
[8]	38,6	900,0	1550 / -
[9]	82	518,0	518 / 1036
Tested SHA1in1	86,1 (11,62 ns)	2203,1	518 / -
Tested SHA2in1	57,3 (17,47 ns)	2931,3	332 / -
Proposed SHA4in1	**33,6** (29,76 ns)	**3441,7**	**212** / -

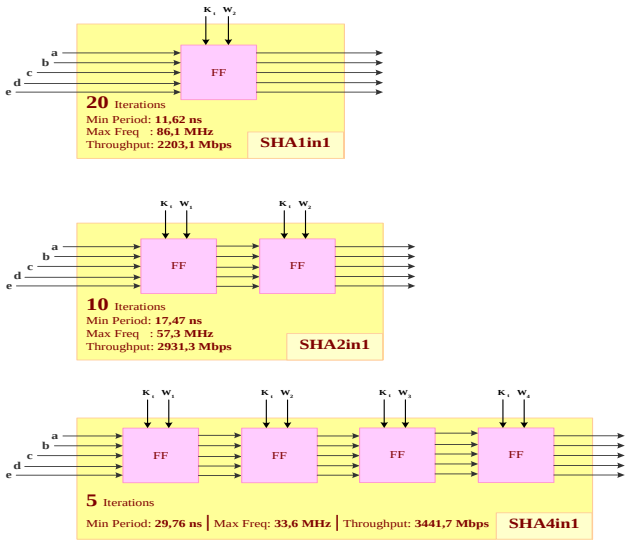


Fig 5*: Comparison of total number of iterations reduced in the three designs namely SHA1in1, SHA2in1 and our proposed solution SHA4in1. The clear reduction in number of iterations and the maximum frequency resulting in low power, high throughout and high speed design is obviously in proposed design.*

$P_{[3]} = 80(P_{op}(f_{op}) + P_{WR}(f_{op}))$
$= 80P_{op}(f_{op}) + 80P_{WR}(f_{op})$
$P_{[4]} = 40(2P_{op}(f'_{op}) + P_{WR}(f'_{op}))$
$= 80P_{op}(f'_{op}) + 40P_{WR}(f'_{op}); \quad f_{op} > f'_{op}$
$P_{[proposed]} = 20(4P_{op}(f''_{op}) + P_{WR}(f''_{op}))$
$= 80P_{op}(f''_{op}) + 20P_{WR}(f''_{op}); \quad f_{op} > f'_{op} > f''_{op}$

VI. *COMPARISON WITH OTHER DESIGNS*

From our test results we can observe the 75% power reduction in the register write/read operation due to reduction in the no. of operation cycles as compared to the [3] and 50% power reduction as compared to [4]. This may introduce 156% increase in the delay of the critical path due to addition of three adder delays in the critical path but this is not significant as this portion resides in the core of each round and does not add significant area penalty.

The operating frequency of the proposed design is dropped significantly from the given throughput from 55 MHz in the [4]. But this operating frequency is bound by the minimum critical path delay of the design that increases by 156% in the proposed design. Therefore this factor has to be considered as well. Even if the operating frequency is kept at 55 MHz then it may suffice for our design as we have already achieved higher throughput.

VII. *CONCLUSION*

A low power and high speed implementation for SHA-1 hash algorithm was presented in this paper that comes out to be more that 15% faster than the previously known designs. In the proposed design we have introduced parallelism to reduce the four iterations into one in each round that takes less time and thus gives higher throughput. Consequently the maximum operating frequency has also reduced contributing to more power saving as operating frequency directly contributes to power consumption.

The proposed implementation has some area penalty that is 156% increase in the critical path delay but that is insignificant due to already small area size covered by the SHA-1 core that keeps the proposed implementation still favorable for mobile and wireless communication devices that are being researched for more low power and high speed solutions.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

REFERENCES

[1] N. Sklavos, P. Kitsos, E. Alexopoulos, and O. Koufopavlou, "*Open Mobile Alliance (OMA) Security Layer: Architecture Implementation and Performance Evaluation of the Integrity Unit*", New Generation Computing: Computing Paradigms and Computational Intelligence, Springer-Verlag, Vol. 23, No 1, pp. 77-100, 2005.

[2] N. Sklavos, E. Alexopoulos, and 0. Koufopavlou, "*Networking Data Integrity: High Speed Architectures and Hardware Implementations*," IAJIT Journal, 1, 0, 54-59, 2003.

[3] A.P. Kakarountas 1, G. Theodoridis 2, T. Laopoulos 2, and C.E. Goutis; "*High-Speed FPGA Implementation of the SHA-1 Hash Function*"; VLSI Design Lab., Elec.&omp. Eng. Dpt., Univ. of Patras, Rio 26110,Greece; Elec. &omp. Lab., Physics Dpt., Ar istotle Univ., Thessaloniki 54124, Greece

[4] Harris Michail, Athanasios P. Kakarountas, Odysseas Koufopavlou, Costas E. Goutis; *"A Low-Power and High-Throughput Implementation of the SHA-1 Hash Function*"; Dpt. of Electrical & Computer Engineering University of Patras, Patras, GR-26500, Greece

[5] S., Dominikus, "*A Hardware Implementation of MD4 Family Hash Algorithms*", in Proceedings of IEEE International Conference on Electronics Circuits and Systems (ICECS'02), Vol. III, pp.1143-1146, Croatia, September 15-18, 2002.

[6] N., Sklavos, G., Dimitroulakos, and O., Koufopavlou, "*An Ultra High Speed Architecture for VLSI Implementation of Hash Functions,",* in Proc.of ICECS, pp. 990-993, 2003.

[7] FIPS PUB 180-1, Secure Hash Standard (SHA-1), National Institute of Standards and Technology (NIST), 2003.

[8] J.M., Diez, S., Bojanic, C., Carreras, and O., Nieto-Taladriz, "*Hash Algorithms for ryptographic Protocols: FPGA Implementations*," in Proc. Of TELEFOR, 2002.

[9] G. Selimis, N. Slavos, and O.Koufopavlou, "*VLSI Implementation of the Keyed-Hash Message Authentication Code For Wireless Application Protocol,"* in Proceedings of ICECS, 900-993, 2003.

[10] J.M. Diez, S. Bojanic, C. Carreras and 0. Nieto-Taladriz, "*Hash Algorithms for Cryptographic Protocols: FPGA Implementations*," in Proceedings of TELEFOR, 2002.

[11] WAP Forum, Wireless Application Protocol, Wireless Transport Layer Security, Architecture Specifications, 2003.

[12] HMAC Standard, The Keyed-Hash Message Authentication Code, National Institute of Standards and Technology (NIST), 2003.

[13] IEEE Std. 801.16-2001, IEEE Standard for Local and Metropolitan Area Networks, part 16, "*Air Interface for Fixed Broadband Wireless Access Systems*," IEEE Press, 2001.

[14] SHA-1 Standard, National Instihite of Standards and Technology (NIST), Secure Hash Standard, FIF'S PUB 180-1, www.itl.nist.gov/tipspuhs/fip80-1.htm

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Various Solutions of Black Hole Attack in A mobile Ad Hoc Network (MANET)

[1] Imad I. Saada [2] Majdi Z. Rashad [3]Sherihan Abuelenin
[1,2,3] Department of Computer Science, Faculty of Computers and Information
Mansoura University, Egypt

***Abstract* - Mobile ad hoc network (MANET) is a kind of wireless network that has a number of nodes, these nodes are distributed and connected without dependency on any infrastructure. MANET security has been an important issue since many years, many researchers have concerned in the black hole threat which "announce it self that it has a route to the destination in all cases". There are many solutions have been proposed to encounter these threats, the problem is that the security threats still exist because it is not prevented or avoided completely, in addition the performance of MANET is adversely affected by these solutions, the objective is to find out to what degree it is possible to prevent this attack by these solutions without causing negative effect on efficiency of MANET, so this survey may facilitate developing or proposing more compact idea to encounter security threats. This paper discusses many important solutions that work to detect a black hole node by using different strategies. In this paper, a new strategy proposed but still under testing.**

Keywords;* *MANET, Black hole, AODV, LIDBPP, Network security.

1. INTRODUCTION

A mobile ad hoc network (MANET) is an autonomous system of mobile nodes, connected by wireless links without existence of any infrastructure, in MANET each node can be considered as a router sending packets for other nodes in the network, these nodes can move continuously and randomly, so there is no fixed topology for MANET. MANET has many challenges such as: the lack of infrastructure, the existence of dynamic topology, power consumption and security threats.

Security threats in MANET presents a larger security challenge if it is compared to conventional wired and wireless networks, mainly due to the common vulnerabilities of wireless connection, one of the most famous security threats in MANET is black hole attack. There are many researches proposed to treat this security problem, some of these researches success in preventing this attack considerably but not completely, the point of comparison or differentiation between these solutions is not only the success degree of the detection, but also is how much these solutions may not affect the performance of the network.

The main interest in this paper is to survey a number of solutions that have detected black hole attacks, and to conclude the differences between them, this process will lead to propose a solution that may collect the advantages of these solutions in one impact solution.

The rest of this paper is organized as follow: Section one discusses some MANET routing protocols, and it will define black hole and its types. Section two has a description of some recent black hole solutions, by analyzing each paper. Section three contains a table which summarizing the important information of each solution. Section four has a new proposed solution called LIDBPP. Section five has the conclusions and the future work.

1.1 MANET routing protocols

The routing protocols of MANET can be divided into three categories [6]:

- Reactive or on-demand routing protocols: example AODV.
- Proactive or table-driven routing protocols: example DSDV.
- Hybrid routing protocols: example ZRP.

1.2 On-Demand Distance Vector routing protocol (AODV)

Most popular reactive routing protocol is AODV, when this routing protocol is implemented in MANET, it will be vulnerable to black hole attacks, most of the black hole attack solutions in this research are applied on AODV routing protocol, so the paper will discuss the algorithm of AODV as follows [17]:

- Each node in MANET has a routing table, To find a route to the destination, the source broadcasts a route request packet (RREQ) immediately to the destination if there is a direct link between source and destination or the source send (RREQ) to the neighboring nodes.
- RREQ contains the destination address, sequence number and broadcast ID.
- Each neighboring node broadcasts (RREQ) to their neighbors, when RREQ reaches an intermediate node, Each node records in its tables the node from which the first RREQ came (this information used for sending RREP).
- If RREQ was previously processed by intermediate node, it will discard duplicate RREQ, by this way the destination or an intermediate node selects the fresher

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

route to the destination based on the destination sequence number, the destination or the intermediate node responds by sending a route reply (RREP) packet to the source node using the path established when the RREQ was sent.

- When the source receives the RREP, it will update its routing table based on the node sending the RREP.
- The source will establish a forward path to the destination, then it will send a data packet to the destination through the path established when the source received the RREP.

1.3 Black holes

[17] Black hole is a malicious node that claims to have the shortest path to the destination, by this way black hole node tricks the source, when the source send a data packet through the black hole to the destination, the black hole absorbs the data packet, so any data packet sent through the black hole will not be reached to the destination.

There are two types of black holes: single black hole attack where MANET just has one black hole node, the second type is a multiple black hole attack where MANET has more than one black hole, these nodes have serious negative effects and the detection process is more complicated than the first type.

2. BLACK HOLE SOLUTIONS

2.1 Detection by checking destination sequence number

Lalit Himral, Vishal Vig and Nagesh Chand [1] proposed a solution for AODV based MANET, it depends on checking the sequence number of source node and intermediate node who has sent back RREP if there is large difference between them or not. Then it will compare the first destination sequence number with the source node sequence number, if there is much more difference between them, then it is clear that the node is a black hole node.

By this solution Black hole node is removed in the initial stage, and because the mechanism is simple and depending just on checking sequence number, there is no modifications on AODV protocol, in addition there is no detection delay nor memory overhead.

2.2 DPRAODV: Solution against black hole attack

Payal N. Raj and Prashant B. Swadas [2] proposed a solution for AODV based MANET, it depends on comparing the RREP sequence number with the threshold value. The threshold value is dynamically updated in every time interval. When the value of RREP sequence number is found to be higher than the threshold value, the solution suspected that the node is a malicious and the node is added to the black list.

The simulation results show that the PDR (packet delivery ratio) increases with same time delay as normal AODV, but the overhead is slightly increased.

2.3 Watchdog solution

Sergio Marti, T.J. Giuli, Kevin Lai, and Mary Baker [3] proposed the watchdog method for DSR based MANET, it detects misbehaving nodes acting alone by maintaining a buffer that contains recently sent packets. When a node forwards a packet, the node's watchdog ensures that the next node in the path also forwards the packet, this means that the node's watchdog listens to the next node. and If the next node does not forward the packet then it is termed as misbehaving.

In this paper two possible extensions are analyzed to DSR to minimize the effects of routing misbehavior in ad hoc networks, the simulation shows that the watchdog increases throughput and overhead in the network.

2.4 Extensions to the watchdog

Animesh Patcha and Amitabh Mishra [4] proposed an extended solution to the watchdog solution for AODV based MANET, the algorithm classifies the nodes into trusted and ordinary nodes. In this algorithm the first assumption is that when a network is formed. The first few nodes that form a network are trusted nodes. Another assumption is that trusted nodes do not show malicious and selfish behavior. Only trusted nodes are selected as watchdogs for a given period of time.

The simulation proves that this algorithm is successful in detecting the presence of colluding malicious nodes, and there is a considerable increase in the network throughput, but it significantly increases the network overhead.

2.5 BDSR Scheme

Po-Chun TSOU, Jian-Ming CHANG, Yi-Hsuan LIN, Han-Chieh CHAO and Jiann-Liang CHEN [5] proposed BDSR scheme for DSR based MANET, it depends on merging proactive and reactive defense architecture in MANET. The BDSR scheme uses a virtual and nonexistent destination address to bait the malicious node to reply RREP. Baited black hole node replies RREP by the above mentioned mechanism. RREP is modified in this scheme to be able to show the address of malicious node, so it is able to detect and block malicious node in the network early.

The results of simulation results show that the packet delivery ratio (PDR) is higher than PDR in case of watchdog solution which uses neighbor node to monitor and detect black hole node, in addition BDSR causes less overhead than watchdog.

2.6 Detection by broadcasting the bluff probe packet (S-ZRP)

Raj Shree, Sanjay Kr. Dwivedi and Ravi Prakash Pandey [6] proposed a solution for ZRP based MANET, in this solution a zone with multiple black hole nodes is considered, the solution implements a Secure-ZRP protocol which can be used to prevent multiple and cooperating black holes attack in MANET, When local communication occurs at that time originator node broadcasts the bluff probe packet, it contains the address of destination but in actual this is the address of nonexistent (virtual) node. This massage is called bluff probe request packet. As it will give response, the source node detects and blocks it as a black hole node. After this, the source node sends information to the direct neighbors for updating their entries.

S-ZRP is an efficient solution to detect the multiple black hole nodes and to stop their attack, the simulation shows how

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

the approach prevent the black hole nodes from receiving and relaying the packets.

2.7 Detection by (DRI) table

Sanjay Ramaswamy, Huirong Fu, Manohar Sreekantaradhya, John Dixon and Kendall Nygard [7] proposed a methodology for AODV based MANET to identify multiple black hole nodes cooperating as a group. AODV routing protocol is slightly modified by this technique, it basically makes use of the Data Routing Information (DRI) table in addition to the cached and current routing tables. In this methodology each node maintains an additional Data Routing Information (DRI) table.

In addition to the identification of multiple and cooperating black hole nodes in a MANET, the secure paths from source to destination can be discovered by avoiding multiple black hole nodes.

2.8 Detection by EDRI table

Gundeep Singh Bindra, Ashish Kapoor, Ashish Narang and Arjun Agrawal [8] proposed a solution for AODV based MANET. The solution is depending on tackling the black hole and gray hole attacks by maintaining an Extended Data Routing Information (EDRI) Table at each node, besides the Routing Table of the AODV protocol. The EDRI table which accommodates the gray behavior of nodes as well, it gives subsequent chances to the nodes identified as black holes.

Detection by EDRI was proposed to be able to deal with multiple cooperating black hole nodes. Detection by EDRI adds some important values to DRI table in [7] to develop the mechanism of detecting the black hole nodes and to take into account the grey behavior of nodes in the network.

2.9 Detection using negotiation with neighbors

Mehdi Medadian and Khossro Fardad [9] proposed a solution for AODV based MANET, in this solution if a node is the first receiver of a RREP packet, this node forwards packets to the source and judges the replier. The judgment process depends on the opinion of network's nodes about replier. The activities of a node are logged by its neighbors, and each neighbor must send their opinion about a node. When all opinions of neighbors are collected by a node, it decides if the replier is a malicious node, the decision is applied according to number of rules.

As in the simulation the proposed solution detects cooperative/multiple black hole nodes and increases performance in terms of packet delivery rate PDR and throughput, which it is better than that of standard AODV, but as appears in the simulation, the proposed solution causes minimal additional delay and overhead.

2.10 Routing security in wireless ad hoc networks (SIDSR)

Deng, H., W. Li and D. Agrawal [10] proposed a solution for AODV based MANET, it uses one more route to the intermediate node that replays the RREQ message, adding one more route is used for checking whether the route from the intermediate node to the destination node exists or not. If it exists, the intermediate node can be trusted. If not, the reply message from the intermediate node can be discarded, the source sends out alarm message to the network and blocks the node from the network.

This paper analyzes one type of attack, the black hole, which can be deployed against a MANET, the authors proposed a feasible solution for AODV based MANET.

2.11 A Local Intrusion Detection Security Routing (LIDSR) mechanism

Maha Abdelhaq, Sami Serhan, Raed Alsaqour and Anton Satria [11] proposed a solution. This solution is called Local Intrusion Detection Security Routing (LIDSR), it is proposed for AODV based MANET. In the LIDSR mechanism the intrusion detection is performed locally using the previous node from the attacker node instead of performing the intrusion detection via the source node, so this solution adds an improvement to the previous solution [10] which is called in this paper (source intrusion detection security routing mechanism SIDSR).

The simulation compares LIDSR with previous solution (SIDSR) [10], it proves that LIDSR causes lower network overhead and time delay, and it increases throughput by changing the number of nodes, network size, and the transmission range.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

3. SUMMARY TABLE

In table-1, some of the solutions approximately have similar methods with some improvements, so the solutions are summarized and exposed in a progressive way, to show how each solution developed the previous solution.

Table 1. A SUMMARY TABLE

Solution / applied on	Method	Black hole attack	Metrics	Findings / compared with
Detection by checking destination sequence number / AODV	checking the destination sequence number	Single	Packet delivery ratio PDR, packet loss	-Increasing PDR -Decreasing packet loss / AODV
DPRAODV / AODV	Comparing RREP sequence number with threshold value	Single	PDR, time delay, overhead	-Increasing PDR -Increasing overhead / AODV
Watchdog solution / DSR	Overhearing the next node in the path by the watchdog to ensure if it forwards the packet or not	single	Throughput, overhead	-Increasing throughput -Increasing overhead / DSR
(Extensions to the watchdog) / AODV	classifying of nodes into trusted nodes and ordinary nodes, the selection of watchdogs from only trusted nodes	multiple	Throughput, overhead	-dealing with multiple cooperating black holes -increasing throughput -increasing overhead / AODV
BDSR / DSR	Sending RREQ with nonexistent destination address to the black hole	Single	PDR, overhead	-Increasing PDR -Decreasing overhead / watchdog solution.
S-ZRP / ZRP	broadcasting the bluff probe packet with nonexistent destination address (detection from the source)	Multiple	Number of query packets that are received and relayed	-dealing with multiple cooperating black holes - detecting and preventing the black holes efficiently / ZRP
Detection by (DRI) table / AODV	Maintaining (DRI) Table at each node with additional routing information	multiple	Not simulated	-dealing with multiple cooperating black holes / AODV
Detection by EDRI table / AODV	Maintaining (EDRI) Table at each node with additional routing information	multiple	Not simulated	-dealing with multiple cooperating black holes -taking into account the grey behavior of nodes / detection by (DRI) table
Detection using negotiation with neighbors / AODV	collecting all opinions of neighbors about the replier	multiple	PDR, Throughput	-dealing with multiple cooperating black holes -increasing PDR and throughput / AODV
Routing security in wireless ad hoc networks (SIDSR) / AODV	The source uses one more route to the intermediate node that replays the RREQ	single	Not simulated	-detecting and avoiding the black hole / AODV
LIDSR / AODV	The previous node uses one more route to the intermediate node that replays the RREQ	Single	Throughput, overhead and time delay	-increasing throughput -decreasing overhead and time delay / SIDSR

4. A PROPOSED SOLUTION: A local Intrusion Detection by Bluff Probe Packet (LIDBPP)

The proposed method is based on bluff packet, it aims to detect and stop the black hole attack in AODV based MANET, this method can deal with multiple black holes attack and will start the detection process by sending a bluff packet that includes a specific virtual destination address, an intermediate node (the previous node from the black hole) sends bluff packet and will take the decision with nonintervention from the source node as follow:

- If the RREQ includes a normal address and the node has a route to the destination it will send RREP.

- If the RREQ includes a normal address and the node has not a route to the destination it will forward RREQ to the next nodes.

- If the RREQ includes the specific virtual address then it is a bluff packet and it must be forwarded, if any node sends this packet and then receives RREP from the next node, it must send block packet because this node is a black hole node.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

- As in figure 1 since a black hole node sends RREP regardless of the address of RREQ, then it will response to bluff packet, so it will be blocked from the previous node.

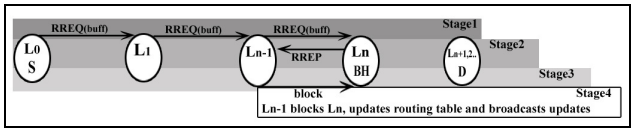


Figure 1. A proposed solution

- After blocking the black hole, the previous node will repeat sending bluff packet to the node that locates next the blocked node and the process will be repeated until blocking all the black hole nodes as in figure 2, there are no need in this process to back to the source node, every intermediate node is responsible to block all black hole nodes that locate next.

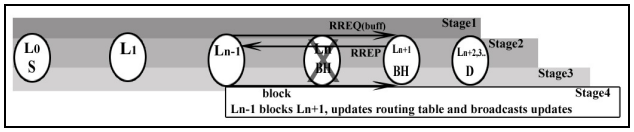


Figure 2.A proposed solution

- Each bluff packet generated from the source will clean the network, because bluff packet is moved from a node to a next node as a serial process.

By starting from the previous node, there is no need to return to the source node, so the detection and blocking process will be occurred with minimal number of packets and in short time, so network overhead and time delay will be minimized, but in S-ZRP [6] we can see that the detection process needs more steps and messages in order to detect and block the black hole node, especially if the distance between the source and black hole node is long, this will negatively affect the network performance such as increasing network overhead and time delay.

The algorithm of LIDBPP is as follow:

L0: source node, L1,2…..n….n+1: intermediate nodes, RREQn: RREQ with normal destination address, RREQs: RREQ with specific and virtual destination address.

Stage1: Source node L0

Generate RREQ

Propagate RREQ

If RREQn Then

Precede normal AODV algorithm

Stage2: Else if RREQs && Ln send RREP to Ln-1 Then

Stage3: Ln-1 send block Ln

Ln receive block

Stage4: Ln-1 updates routing table and broadcasts updates

Else

Ln sends RREQs to Ln+1

End if

We try to propose a new method to detect a black hole node by overcome the most drawbacks of the previous works. This proposed method now is under evaluations.

5. CONCLUSION AND FUTURE WORK

This paper introduced many recent solutions that work to detect a black hole node by using different methods, and it explained how these methods worked, the paper has included a summary table that contains the analyzed information of each solution. It was obvious in the paper that the process of developing a method with high efficiency and less negative impact on the performance of a network is a challenge, so the development effort is an active process.

This paper can be used as a reference to know where researchers arrived in the issue of MANET security, this may support the development process of the solutions with keeping the performance of MANET.

The paper also included a new solution to detect multiple black holes based on bluff probe packet (contains a specific virtual destination address) that tricks the black hole.

The process of developing this new method would detect and block the black hole nodes with high efficiency and less negative impact on MANET performance. In the future, the new solution (LIDBPP) must be simulated to evaluate the performance of MANET and to compare it with the other solutions.

REFERENCES

[1] Lalit Himral, Vishal Vig, Nagesh Chand, (2011) "Preventing AODV Routing Protocol from Black Hole Attack", International Journal of Engineering Science and Technology (IJEST), Vol. 3 No. 5.

[2] Payal N. Raj, Prashant B. Swadas, (2009) "DPRAODV: A dynamic Learning System Against Blackhole Attack in AODV Based MANET", IJCSI International Journal of Computer Science Issues, Vol. 2.

[3] S. Marti, T.J. Giuli, K. Lai, M. Baker, "Mitigating routing misbehavior in mobile ad hoc networks," 6th MobiCom, Boston, Massachusetts, August 2000

[4] Animesh Patcha and Amitabh Mishra, (2003) "Collaborative Security Architecture for Black Hole Attack Prevention in Mobile Ad Hoc Networks", IEEE

[5] Po-Chun TSOU, Jian-Ming CHANG, Yi-Hsuan LIN, Han-Chieh CHAO, Jiann-Liang CHEN, (2011) "Developing a BDSR Scheme to Avoid Black Hole Attack Based on Proactive and Reactive Architecture in MANETs", ICACT2011.

[6] Raj Shree, Sanjay Kr. Dwivedi, Ravi Prakash Pandey, (2011) "Design Enhancements in ZRP for Detecting Multiple Black Hole Nodes in Mobile Ad Hoc Networks", International Journal of Computer.

[7] S. Ramaswamy, H. Fu, M. Sreekantaradhya, J. Dixon, and K. Nygard, (2003) "Prevention of cooperative black hole attack in wireless ad hoc networks", 2003 International Conference on Wireless Networks (ICWN'03), pages 570–575. Las Vegas, Nevada, USA.

[8] Gundeep Singh Bindra, Ashish Kapoor, Ashish Narang, Arjun Agrawal, (2012) "Detection and Removal of Co-operative Blackhole and Grayhole Attacks in MANET", 2012 International Conference on System Engineering and Technology, Bandung, Indonesia.

[9] Mehdi Medadian, Khossro Fardad, (2012) "Proposing a Method to Detect Black Hole Attacks in AODV Routing Protocol", European Journal of Scientific Research.

[10] Deng, H., W. Li and D. Agrawal, 2002. "Routing security in wireless ad hoc networks". IEEE communications magazine, 40(10): 70-75.

[11] Maha Abdelhaq, Sami Serhan, Raed Alsaqour, Anton Satria, (2011) "Security Routing Mechanism for Black Hole Attack over AODV MANET Routing Protocol", Australian Journal of Basic and Applied Sciences.

[12] Panagiotis Papadimitratos, Zygmunt J. Haas, (2003) "Secure Data Transmission in Mobile Ad Hoc Networks", ACM Workshop on Wireless Security (WiSe 2003), San Diego.

[13] Chandni Garg, Preeti Sharma, Prashant Rewagad, (2012) "A Literature Survey of Black Hole Attack on AODV Routing Protocol", International Journal of advancement in electronics and computer engineering (IJAECE).

[14] Akanksha Saini, Harish Kumar, (2010) "Comparison between Various Black hole Detection Techniques in MANET", NCCI 2010 -National Conference on Computational Instrumentation, India.

[15] Panagiotis Papadimitratos, Zygmunt J. Haas, (2002) "Secure Routing for Mobile Ad hoc Networks", the SCS Communication Networks and Distributed Systems Modeling and Simulation Conference (CNDS).

[16] Peter Sholander, Andreas Yankopolus and Paul Coccoli (2002) "Experimental comparison of hybrid and proactive MANET routing protocols", MILCOM 2002, IEEE.

[17] Satoshi K., Hidehisa N., Nei K., Abbas J., and Yoshiaki N., (2007) "Detecting Blackhole Attack on AODV-based Mobile Ad Hoc Networks by Dynamic Learning Method", International Journal of Network Security.

AUTHORS PROFILE

Imad I. Saada is a PHD student in computer science department in Mansoura University and a member of the academic staff at IT. department in AL-Quds Open University. His subject is in the distributed systems.

Magdy Z. Rashad is a professor at computer science department in Mansoura University. He is the decision support systems unit coordinator at faculty of computers &information in Mansoura University. He has supervised over 10 PhDs and 21 masters mostly specialized in artificial intelligence and its applications related to real life. As a result of his work he has published over 84 papers. Current project is grid computing.

Sherihan M.Abuelenin is an assistant professor at computer science department in Mansoura University. She received the Bachelor of Computer Science degree from Mansoura University, Mansoura, Egypt, in 2002, the Ph.D. degree from Chiba University, Chiba, Japan, in 2012. Her interest works are cloud computing, grid systems, security, and internet of things.

Tools and Techniques for Ontology Interoperability: A Survey

R. Lakshmi Tulasi
Professor& HOD, Department of IT, QISCET,
Ongole, India

Dr M. Srinivasa Rao
Professor, Dean CIHL, SIT, JNTUH,
Hyderabad, India

Abstract— The idea of the semantic web is to add machine process able information to web-based data in order to realize interoperability. Ontology is a shared conceptualization of knowledge representation of particular domain. These are used for the enhancement of semantic information explicitly. Ontologies play a prominent role in the concept of the semantic web to provide semantic information for assisting communication among heterogeneous information repositories. Ontology Interoperability provides the reusability of ontologies Different domain experts and ontology engineers create different ontologies for the same or similar domain depending on their data modeling requirements. These cause ontology heterogeneity and inconsistency problems. As increasing numbers of ontologies are developed by diverse communities, the demand for rapid ontology mapping is arising. For more better and precise results ontology mapping is the solution. As their use has increased, providing means of resolving semantic differences has also become very important. Papers on ontology interoperability report the results on different frameworks and this makes their comparison almost impossible. Therefore, the main focus of this paper will be on providing some basics of ontology interoperability and briefly introducing its different approaches. In this paper we survey the approaches that have been proposed *for providing interoperability among domain ontologies and* its related techniques and tools.

Keywords- *Ontology Interoperability; Ontology Mapping; Ontology Alignment; Ontology Merging; Semantic heterogeneity; Semantic web;*

I. INTRODUCTION

The WWW has become a vast resource of information. It is growing rapidly from last few decades. The problem is that finding the information, and the individual desires are often quite difficult, because of complexity in organization and quantity of the information stored. In traditional search engines, Information Retrieval (IR) is keyword based or with a natural language. Query entered by the users is not understandable, so it retrieves the large number of documents in the ranked order which have poor semantic relationships among the documents. This keyword based approach results poor precision - List of retrieved documents contain a high percentage of irrelevant documents, and poor recall- List of relevant retrieved among possible relevant. To avoid the above problems semantic search engines are required.

Ontology is used to model knowledge representation of a particular domain (E-learning, sports, medical, etc). Ontologies are explicit specifications of the conceptualization and corresponding vocabulary used (Gruber 1993). Ontology is the fundamental factor for semantic web. So, users create different ontologies depending on their data modeling requirements for the same or similar domain. They are free to use vocabulary of their own. This leads to heterogeneity and inconsistency problems.

The basic operation we perform to solve above problems among ontologies is "mapping" which interprets the sets of correspondences between similar concepts and among two or more ontologies of same or similar domains. This is prominent research area in the field of AI (Artificial Intelligence). These mappings support two other related operations ontology alignment and ontology merging. Ontology alignment process takes two or more input ontologies and produces a set of relationships between concepts that match semantically with each other. These matches are also called mappings. Ontology merging, as its name implies merges two ontologies of same or similar domain in to one based on semantic similarity of concepts and produces unique ontology. Three important mismatches may exist between ontologies syntactic, semantic and lexical mismatches. Our recent researchers developed several methods and techniques to identify these mismatches.

The rest of the paper organized as follows. Section II discusses about different types of ontology interoperability, Section III discusses about types of ontology mapping. Section IV discusses about challenges in ontology mapping. Section V discusses about types of mismatches. Section VI discusses about tools and techniques used for ontology interoperability.

II. ONTOLOGY INTEROPERABILITY

This section describes several operations on ontologies like Transformation and translation, merging, mapping, Integration. These can be considered as an ontology reuse process. [1, 2]

A. OntologyTransformation and Translation

Ontology Transformation [14, 15] is the process used to develop a new ontology to cope with new requirements made

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

by an existing one for a new purpose, by using a transformation function't'. Many changes are possible in this operation, including changes in the semantics of the ontology and changes in the representation formalism. Ontology Translation is the function of translating the representation formalism of ontology while keeping the same semantic. In other words, it is the process of change or modification of the structure of ontology in order to make it suitable for purposes other than the original one. There are two types of translation. The first is translation from one formal language to another, for example from RDFS to OWL, called syntactic translation. The second is translation of vocabularies, called semantic translation [14]. The translation problem arises when two Web-based agents attempt to exchange information, describing it using different ontologies.

B. Ontology Merging

Ontology merging [7, 10, 15] is the process of creating a new single coherent ontology from two or more existing source ontologies related to the same domain. The new ontology will replace the source ontologies.

C. Ontology Integration

Integration [7, 10] is the process of creating a new ontology from two or more source ontologies from different domains.

D. Ontology Alignment

Ontology alignment [3,4,5,6] is the process or method of creating a consistent and coherent link between two or more ontologies by bringing them into mutual agreement. This method is near to artificial intelligence methods: being a logical relation, ontology alignments are used to clearly describe how the concepts in the different ontologies are logically related. This means that additional axioms describe the relationship between the concepts in different ontologies without changing the meaning in the original ontologies. In fact the ontology alignment uses as a pre process for ontology merging and ontology integration. There are many different definitions for ontology alignment depending upon its applications and its intended outcome.

Sample definitions include the following

- Ontology alignment is used to establish correspondences among the source ontologies, and to determine the set of overlapping concepts, concepts that are similar in meaning but have different names or structure, and concepts that are unique to each of the sources [15].
- Ontology alignment is the process of bringing two or more ontologies into mutual agreement, making them consistent and coherent.

Given two ontologies O1 and O2, mapping of one ontology in to another means that each entity (concept c, relation R, Instance I) in ontology is trying to find a corresponding entity which has the same intended meaning in ontology O2.

Formally, an ontology alignment function is defined as follows:

An ontology alignment function, align based on the set E of all entities e € E and based on the set of possible ontologies O, is a partial function.

Align: O1 → O2
Align (eO1) = fO2 if Sim (eO1, fO2) > threshold.
Where Oi: ontology, eOi, fOj: entities of (Oi, Oj,)
Sim (eO1, fO2): Similarities function between two entities eO1 and fO2.

The ontology alignment function is based on different similarit y measures.
A similarity measure is a real valued function Sim(ei, fj): O×O → [0, 1] measuring the degree of similarity between x and y.
Ontology heterogeneity is shown in Fig 1.

E. Ontology Mapping

Ontology mapping [6, 9, 14, 12, 13] is a formal expression or process that defines the semantic relationships between entities from different ontologies. In other words, it is an important operator in many ontology application domains, such as the Semantic Web and e-commerce, which are used to describe how to connect and from correspondences between entities across different ontologies. Ontology matching is the process of discovering similarities between two ontologies. An entity 'e' is understood in an ontology O denoted by e|O is concept C, relation R, or instance I, i.e. e|O € C U R U I. Mapping the two ontologies, O1 onto O2, means that each entity in ontology O1 is trying to find a corresponding entity which has the same intended meaning in ontology O2.

The Ontology mapping function "map" is defined based on the vocabulary, E, of all terms e € E and based on the set of possible ontologies, O as a partial function:

$$\text{map: } E \times O \times O \rightarrow E, \text{ with}$$
$$e \in O1(\exists\ f \in O2 : \text{map}(e,O1,O2) = f \vee \text{map}(e,O1,O2) = \perp).$$

An entity is mapped to another entity or none.

III. Types of Ontology Mapping

Based on the method of ontology mapping and how ontologies are created and maintained, it is divided in to three categories.

A.Ontology mapping between an integrated global ontology and local ontologies.[16,17]:

In this case, ontology mapping is used to map a concept of one ontology into a view, or a query over other ontologies.

B.Ontology mapping between local ontologies [25]:

In this case, ontology mapping is the process that transforms the source ontology entities into the target ontology entities based on semantic relation. The source and target are semantically related at a conceptual level.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

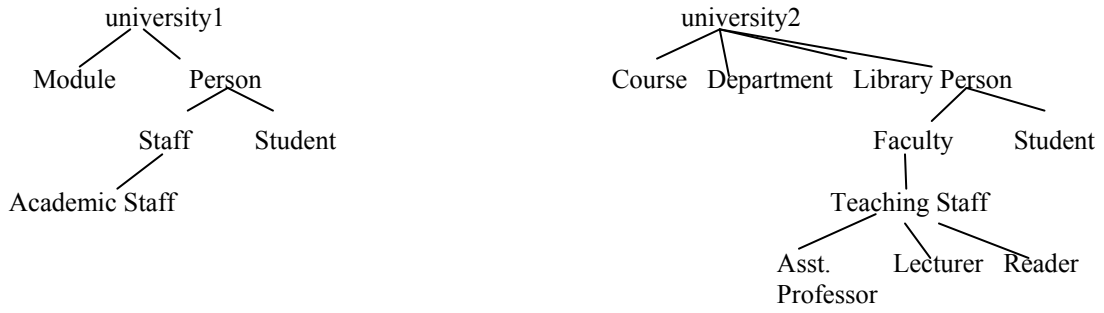


Figure 1.Ontology heterogeneity among ontologies of same domain

C. Ontology mapping in ontology merge and alignment[15]:

In this case, ontology mapping establishes correspondence among source (local) ontologies to be merged or aligned, and determines the set of overlapping concepts, synonyms, or unique concepts to that sources[15]. This mapping identifies similarities and conflicts between the various source (local) ontologies to be merged or aligned.

IV. Challenges of Ontology Mapping

In this section, we discuss challenges of ontology mapping

1. Large-scale evaluation
2. Performance of ontology-matching techniques
3. Discovering missing background knowledge
4. Uncertainty in ontology matching
5. Matcher selection and self-configuration
6. User involvement
7. Explanation of matching results
8. Social and collaborative ontology matching
9. Alignment management: infrastructure and support
10. Reasoning with alignments.

V. Types of Mismatches

Different types of mismatches may occur between different ontologies. Indeed different ontology designers opt for different representation languages and use different ontology editors to represent knowledge at different levels of granularity (detail). This explains the emergence of different forms of ontology mismatches. The identification of these types of mismatches is essential in order to solve them during the mapping, alignment or merging process.

A. Syntactic mismatches

Two ontologies are syntactically heterogeneous if they are represented by different representation languages, such as OWL, KIF etc. To resolve this type of mismatches, simply transform the representation language of one ontology to the representation language of the other ontology. Herein, we state that sometimes the translation is difficult and even impossible.

B. Lexical mismatches

Describe the heterogeneities among the names of entities, instances, properties, or relations. In this type of mismatches, we may find four forms of heterogeneities: Synonyms, Homonyms, Same name in different languages, and same entities with the same name but with different syntactic variations.

C. Semantic mismatches

These kind of mismatches describe words belong to same synonym set. For example, ontology A has price and ontology B has cost. Then both are said to be semantically equivalent or match, otherwise it is a mismatched pair.

VI. Tools and Techniques for Ontology Operations

A. LSD[16] (Learning Source Description):

LSD semi automatically creates semantic mappings with a multi strategy learning approach. This approach employs multiple learner modules with base learners and the meta-learner where each module exploits a different type of information in the source schemas or data. LSD uses the following base learners: 1) The Name Learner: it matches an XML element using its tag name, 2) The Content Learner: it matches an XML element using its data value and works well on textual elements, 3) Naïve Bayes Learner: it examines the data value of the instance, and doesn't work for short or numeric fields, and 4) The XML Learner: it handles the hierarchical structure of input instances. Multi-strategy learning has two phases: training and matching. In the training phase, a small set of data sources has been manually mapped to the mediated schema and is utilized to train the base learners and the Meta learner. In the matching phase, the trained learners predict mappings for new sources and match the schema of the new input source to the mediated schema.

B. MOMIS[17](Mediator Environment for Multiple Information Sources):

MOMIS creates a global virtual view (GVV) of information sources, independent of their location or their data's heterogeneity. MOMIS builds an ontology through five phases as follows:

1) Extraction of local schema

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

2) Local source annotation using Word Net (online dictionary)

3) Common thesaurus generation: relationships of inter-schema and intra-schema knowledge about classes and attributes of the source schemas

4) Generation of GVV: A global schema and mappings between the global attributes of the global schema and source schema are generated.

5) GVV annotation is generated by exploiting annotated local schemas and mappings between local schemas and a global schema.

C. A Framework for OIS [20] (Ontology Integration System):

Mappings between an integrated global ontology and local ontologies are expressed as queries and ontology as Description Logic. Two approaches for mappings are proposed as follows: 1) concepts of the global ontology are mapped into queries over the local ontologies (global-centric approach), and 2) concepts of the local ontologies are mapped to queries over the global ontology (local centric approach).

D. GLUE[21]:

It semi-automatically creates ontology mapping using machine learning techniques. It consists of Distribution

Estimator, Similarity Estimator, and Relaxation Labeler. It finds the most similar concepts between two ontologies and by using a multi-strategy learning approach calculates the joint probability distribution of the concept for similarity measurement. It has Content Learner, Name Learner, and Meta Learner. Content and Name Learners are two base learners, while Meta Learner combines the two base learners' prediction. The Content Learner exploits the frequencies of words in content of an instance and uses the Naïve Bayes' theorem. The Name Learner uses the full name of the input instance. The Meta-Learner combines the predictions of base learners and assigns weights to base learners based on how much it trusts that learner's predictions.

E. ONION[22] (ONtology compositION system):

It resolves terminological heterogeneity in ontologies and produces articulation rules for mappings. The linguistic matcher identifies all possible pairs of terms in ontologies and assigns a similarity score to each pair. If the similarity score is above the threshold, then the match is accepted and an articulation rule is generated. After the matches generated by a linguistic matcher are available, a structure-based matcher looks for further matches. An inference-based matcher generates matches based on rules available with ontologies or any seed rules provided by experts. Multiple iterations are required for generating semantic matches between ontologies. A human expert chooses, deletes, or modifies suggested matches using a GUI tool.

F. LOM[23] (Lexicon-based Ontology Mapping):

LOM finds the morphism between vocabularies in order to reduce human labor in ontology mapping using four methods: whole term, word constituent, synset, and type matching. LOM does not guarantee accuracy or correctness in mappings and has limitations in dealing with abstract symbols or codes in chemistry, mathematics, or medicine.

G. QOM[24] (Quick Ontology Mapping):

QOM is an efficient method for identifying mappings between two ontologies because it has lower run-time complexity. In order to lower run-time complexity, light weight ontologies QOM uses a dynamic programming approach. A dynamic programming approach has data structures which investigate the candidate mappings, classify the candidate mappings into promising and less promising pairs, and discard some of them entirely to gain efficiency. It allows for the ad-hoc mapping of large size, light-weight ontologies.

H. PROMPT[27]

PROMPT is a semi-automatic ontology merging and alignment tool. It begins with the linguistic-similarity matches for the initial comparison, but generates a list of suggestions for the user based on linguistic and structural knowledge and then points the user to possible effects of these changes.

I. Onto Morph[28]

Onto Morph provides a powerful rule language for specifying mappings, and facilitates ontology merging and the rapid generation of knowledge-base translators. It combines two powerful mechanisms for knowledge-base transformations such as syntactic rewriting and semantic rewriting. Syntactic rewriting is done through pattern-directed rewrite rules for sentence-level transformation based on pattern matching. Semantic rewriting is done through semantic models and logical inference.

J. Anchor-PROMPT[25]

Anchor-PROMPT takes a set of anchors (pairs of related terms) from the source ontologies and traverses the paths between the anchors in the source ontologies. It compares the terms along these paths to identify similar terms and generates a set of new pairs of semantically similar terms.

K. CMS[26] (CROSI Mapping System)

CMS is an ontology alignment system. It is a structure matching system on the rich semantics of the OWL constructs. Its modular architecture allows the system to consult external linguistic resources and consists of feature

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

generation, feature selection, multi-strategy similarity aggregator, and similarity evaluator.

L. FCA-Merge[29]

FCA-Merge is a method for ontology merging based on Ganter and Wille's formal concept analysis [28], lattice exploration, and instances of ontologies to be merged. The overall process of ontology merging consists of three steps: 1) instance extraction and generation of the formal context for each ontology, 2) the computation of the pruned concept lattice by algorithm TITANIC29, and 3) the non automatic generation of the merged ontology with human interaction based on the concept lattice.

M. CHIMAERA[30]

CHIMAERA is an interactive ontology merging tool based on the Ontolingual ontology editor. It makes users affect merging process at any point during merge process, analyzes ontologies to be merged, and if linguistic matches are found, the merge is processed automatically, otherwise, further action can be made by the user. It uses subclass and super class relationship.

N. *ConcepTool [31]*

This is an interactive and analysis tool that aims to facilitate knowledge sharing. It supports ontology alignment process where the ontologies are represented in Entity Relationship model resulting from reasoning based on description logic. ConcepTool is based on heuristic and linguistic inferences to compare attributes of two entities belonging to the input ontologies. The analyst is then charged of identifying relevant information to resolve conflicts between overlapping entities. Overlapping entities are related to each other through semantic bridges. Each bridge provides a semantic transformation rule to solve the semantic mismatches between these entities. Summarizing, ConcepTool begins by analyzing the input models to derive taxonomic links and overlapping entities. Then, the analyst matches the common entities. The articulation ontology entities are automatically generated and the analyst defines mappings between the attributes of the matched entities. Finally, the articulation ontology is analyzed.

7. CONCLUSION

The ontology Interoperability is a prominent issue in many application domains such as semantic query processing, data integration, data-warehousing, E-Commerce and E-Business. Issues of heterogeneity and inconsistency among the ontologies of same or similar domains will be resolved using ontology mapping. Definitions of ontology matching, ontology merging, ontology Integration are given. We have presented a general framework situating ontology Mapping. Kinds of ontology mapping are proposed. Ten challenges which we face while mapping ontologies are presented. We have located three forms of mismatches that are usually studied in these processes, namely, lexical, syntactic and semantic mismatches.

Because of the wide usage of ontology Interoperability techniques there is a need to consolidate different techniques and tools have been proposed to handle ontology Alignment, ontology Mapping and Merging processes. In this paper, we have surveyed the literature of these techniques and described the different criteria and approaches adopted by algorithms.

REFERENCES

[1] Yannis Kalfoglou, Marco Schorelmmer, "Ontology Mapping: The State of the Art", The Knowledge Engineering Review, Vol. 18:1, 1-31, 2003.

[2] .Helena Sofia Pinto, Joao P. Martins, "A Methodology for Ontology Integration", Proceedings of the International Conference on Knowledge Capture,Technical papers, ACM Press, pp. 131-138, 2001.

[3] .M. Ehrig and J. Euzenat, "State of the Art on Ontology Alignment", Knowledge Web Deliverable D2.2.3, University of Karlsruhe, 2004.

[4]. J. Euzenat and P. Shvaiko, "Ontology Matching", Springer-Verlag, Heidelberg (DE), 2007.

[5]. J. Euzenat and P. Valtchev, "Similarity-Based Ontology Alignment in OWL-Lite", In Proceedings of ECAI, 2004, pp.333-337

[6]. F. Giunchiglia, P. Shvaiko, and M. Yatskevich, "Semantic Schema Matching", In Proceedings of OTM Conferences (1), 2005, pp.347-365.

[7]. C. Ghidini and F. Giunchiglia, "A Semantics for Abstraction", In Proceedings of ECAI, 2004, pp.343-347.

[8]. O. Gotoh, "An Improved Algorithm for Matching Biological Sequences", Presented at Journal of Molecular Biology, 162:705-708, 1982.

[9]. Y. Kalfoglou and W.M. Schorlemmer, "IF-Map: An Ontology-Mapping Method Based on Information-Flow Theory", Presented at Journal Data Semantics, 2003, pp.98-127.

[10]. M.C.A. Klein and D. Fensel, "Ontology Versioning on the Semantic Web", In Proceedings of SWWS, 2001, pp.75-91.

[11]. N.F. Noy and M.A. Musen, "PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment", In Proceedings of AAAI/IAAI, 2000, pp.450-455.

[12]. E. Rahm, P.A. Bernstein, "A Survey of E. Rahm, P.A. Bernstein, "A Survey of Approaches to Automatic Schema Matching", Presented at VLDB Journal, 2001, pp.334–350.

[13]. P. shivaiko and J.Euzenat " A survey of schema based mapping approaches ", presented at Journal of Data Semantics IV 2005, pp. 146-171

[14]. H. Chalupsky, "OntoMorph: A Translation System for Symbolic Knowledge", In Proceedings of KR, 2000, pp.471-482.

[15].D. Dou, D. McDermott, and P. Qi, "Ontology Translation on the Semantic Web", Presented at on Data Semantics Journal, 3360:35–57, 2005.

[16]. AnHai Doan, Pedro Domingos, Alon Halevy, "Learning to Match the Schemas of Data Sources: A Multistrategy Approach", Machine Learning, 50 (3): 279- 301, March 2003.

[17].Domenico Beneventano, Sonia Bergamaschi,Francesco Guerra, Maurizio, "Synthesizing an Integrated Ontology", IEEE Internet Computing, September - October 2003.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[18]. N. Noy and M. Musen, "PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment", Proceedings of the National Conference on Artificial Intelligence (AAAI), 2000.

[19]. Nuno Silva and Joao Rocha, "Ontology Mappingfor Interoperability in Semantic Web", Proceedings of the IADIS International Conference WWW/Internet 2003 (ICWI'2003). Algarve, Portugal; November 2003.

[20]. Calvanese, D, De Giacomo, G and Lenzerini, M,2001a, "A Framework for Ontology Integration", Proceedings of the 1st International Semantic Web Working Symposium (SWWS) 303–317.

[21]. AnHai Doan, Jayant Madhavan, Pedro Domingos, Alon Halevy, "Learning to Map between Ontologies on the Semantic Web", VLDB Journal, Special Issue on the Semantic Web, 2003.

[22]. Mitra, P and Wiederhold, G, "Resolving Terminological Heterogeneity in Ontologies", Proceedings of the ECAI'02 workshop on Ontologies and Semantic Interoperability, 2002.

[23]. John Li, "LOM: A Lexicon-based Ontology Mapping Tool", Proceedings of the Performance Metrics for Intelligent Systems (PerMIS. '04), 2004.

[24].Marc Ehrig, Steffen Staab, "QOM – Quick OntologyMapping", GI Jahrestagung (1), 2004.

[25]. N. Noy and M. Musen, "Anchor-PROMPT: Using Non-Local Context for Semantic Matching", Proceedings of the Workshop on Ontologies and Information Sharing at the International Joint Conference on Artificial Intelligence (IJCAI), 2001.

[26]. Yannis Kalfoglou, Bo Hu, "CROSI Mapping System (CMS) Results of the 2005 Ontology Alignment Contest",

K-CAP Integrating Ontologies Workshop 2005, Banff, Alberta, Canada, 2005.

[27]. N. Noy, and M. Musen, "PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment" Proceedings of the National Conference on Artificial Intelligence (AAAI), 2000.

[28]. H. Chalupsky. "Ontomorph: A Translation System for Symbolic Knowledge", Principles of Knowledge Representation and Reasoning, 2000.

[29].Gerd Stumme, Alexander Maedche, "FCA-Merge: Bottom-Up Merging of Ontologies", In proceeding of the International Joint Conference on Artificial Intelligence IJCA101, Seattle, USA, 2001.

[30].D. McGuinness, R. Fikes, J. Rice, and S. Wilder, "The Chimaera Ontology Environment", In Proceedings of the 17th National Conference on Artificial Intelligence (AAAI), 2000.

[31]. E.Compatangelo ,and H. Meisel, Intelligent support to knowledge sharing through the articulation of class schemas", proceedings of the 6th International Conference on Knowledge-Based Intelligent Information and Engeneering Systems, Italy, 2002.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Generic Lightweight Certificate Management Protocol (GLCMP)

Shahzad Khan
Department of Computer Science
Shaheed Benazir Bhutto University(SBBU)
Khyber PukhtunKhwa, Pakistan

Muhammad Asif
Department of Communications System Engineering
School of Electrical Engineering and Computer Sciences(SEECS), NUST, Islamabad, Pakistan

Abstract-*This paper describes a Generic Light Weight Certificate Management Protocol (GLCMP) for handling certificates on mobile devices. Theoretically, various security solutions are designed to protect the valuable information of mobile users. But, its power, memory and processing constraints, high response time and authentication latencies are the main challenges for the researcher to develop and integrate standard security mechanisms in it. It is observed that, most of mobile users are not technical enough to configure security parameters and even already developed libraries do not support extended security features like transparent handling of certificates, verification of identities, and distribution of certificates. In this paper, an innovative and comparatively efficient protocol is designed and implemented. It does not only overcome the shortcoming of the certificate handling in mobile devices but also provides some extended certificate related features like registration, authentication and trust delegation. The designed GLCMP is lightweight because all complex and computation-intensive operations, involved in creation of certificate request in PKCS#10 standard format, are offloaded to a proxy server. It also provides domain based secure registration and verification of the identities without exchanging any confidential information to the proxy servers and even no user's credential is exchanged on network for authentication. After analyzing its performance, we noticed that authentication latency of GLCMP is 0.394 sec which is less than previously proposed protocols like NSI (4.7), PKI (5.01), and PKASSO (5.19 delegation time + 0.082 authentication times). We also formally verified our designed by using Z-Notation Modeling techniques and found that it is protected against man-in-the-middle, replay and impersonation and non-repudiation attacks.*

I. INTRODUCTION

Mobile devices and digital gadgets are very popular and commonly used in daily life. The use of these gadgets has also started in business and e-commerce. Research community increased its processing power and designed new advanced applications to attract business community but still business community have security concerns like authentication, authorization, integrity, confidentiality and non-repudiation. Even there are many problems in the development of secure applications for mobile devices. First, most of the users are not technical enough to configure security parameters and even already developed libraries do not support extended security features like transparent handling of certificates, verification of identities, and distribution of certificates. It is also observed that already developed security libraries are very difficult to use and integrate with existing applications to provide security features.

Public key Infrastructure (PKI) [1] is the most reliable mechanism for achieving end-to-end security in desktop environment. But PKI is not feasible in mobile devices due to memory, battery and processing speed constraints. A lot of protocols have been proposed for authentication of mobile device like Kerberos, PKI, PKINIT [3], SaPKI [4], NSI [12], WPKI [2] and PKASSO but they are not suitable for mobile devices because of some protocols do not provide non-repudiation and signature like Kerberos, PKINIT, MP-PKINIT [19], some are very complex like PKASSO and authentication latency of some protocols is not affordable like NSI, PKI.

In order to solve above problems, a Generic Light Weight Certificate Management Protocol (GLCMP) is designed which is based on holistic approach in order to solve complex certificate management tasks. GLCMP transparently verifies users, generates, certifies and manages certificates for mobile devices based on well-established standards i.e. PKCS#10[5], and PKCS#7[11]. Salient features of this protocol are following.

- It is light weight because it offloads all the cryptographic computational intensive operations, involved in creating certificate request in PPKCS#10 standard, to a proxy server.
- Trust between mobile device and proxy server is developed without exchanging any secret information on network.
- Provides secure registration and identity verification.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

- Provides suitable authentication latency for mobile devices.
- Provides authentication, integrity and non-repudiation.
- Based on generic security objects that are easy to extend, use and integrate.

II. RELATED WORK

Various security mechanisms have been proposed to authenticate mobile users by using existing network authentication protocols like PKI, PKINT or proposing different protocols.

Fang LIU, Qi Lang proposed WPKI protocol. In this protocol, WAP gateway is used to server authentication services and build trust channel between mobile terminal and bank server. Mobile terminal uses WTLS secure session to communicate with WAP gateway. The mobile device must have WIM module (Wireless identity module: contain User's private key and User's certificate URL directory) for requesting transfer of money. User will be prompted to select certificate, enter PIN and then transmit certificate URL to WAP gateway. To solve the mobile node's issue of searching and verifying digital certificates, Jalali-Sohi and P. Ebinger presented PKI-server based authentication infrastructure [12]. Mobile node delegates its responsibility to PKI-Server. In this method, minimum achieved authentication latency is 4.75sec which is not suitable.

So, Ki-Woong park et al, have presented a security infrastructure called PKASSO which consists of five main components. PKINIT is one of the main components, enhanced form of Kerberos, which encompasses Kerberos, LDAP and CA servers. A service requesting entity, mobile user, for authentication, authorization and accounting should have smart card like device called PANDA equipped with low powered Zigbee for intercommunication and location sensing. Third component is a service device with Zigbee for communication with the users. Fourth component is a delegation server for maintain all the proxy certificates along with public private keys delegated and signed by users. In start user delegates its all authentication process to delegation server according to RFC3820 [7]. The last and the most important component is referee server which is assigned the duty of investigating authentication messages and binding these messages with users to provide non-repudiation. User performs mutual authentication by using PKI and then delegate other authentication operation to delegation server. For this, User create proxy certificate with public key sent by Delegation server, sign it by its private key and send to Delegation server. After successful delegation, user encrypt received challenge message from service device with AES twice and send it to Delegation server. Delegation server verifies this message from referee server and after successful verification performs PKINIT operations. In PKINIT operations, delegation server gets TGT, Ticket Granting Ticket, by PKI and SGT over Kerberos. This protocol achieve, Single Sign-on, Digital Signature, Authentication, Non-repudiation and secure key distribution on the cost of 0.082 sec authentication latency excluding 5.19 sec spent for delegation.

Liang Cai, Xiaohu Yang and Chun Chen presents a SaPKI architecture to offload cryptographic computation-intensive operations involved in creation of RSA keys for generating digital signature over GSM and CDMA network. In SaPKI, Modadugu's protocol [13] for key exchange and Asokan's S^3 [14] for generating signature are used to achieve efficiency. SaPKI is implemented in service provider's premises and provides three interfaces for offloading. ISaPKI_KeyGen ( ) helps to generate keys for encryption and key exchange, ISaPKI_Cert ( ) initialize the keys and ISaPKI_Sign ( ) for signing the message. Further, to minimizes computation cost and overhead, Mohsen Toorani and Ali Asghar proposed a secure infrastructure called LPKI for computation-constrained platforms like mobile devices and digital gadgets [8]. It uses Elliptic Curve Cryptography (ECC) and signcryption. The *Elliptic Curve Cryptography* (ECC) is usually deemed as a suitable solution for the resource-constrained devices [16]. As an example, it is believed that a 160-bit key in an elliptic curve-based system provides the same level of security as that of a 1024-bit key in an RSA-based system [17]. For exchanging keys, LPKI takes the advantages of HMQV key exchange protocol [19] because of its efficiency, standardization and a lot of provided security attributes. So enable efficient certificate management in mobile phones, five previously proposed security mechanisms have been presented. In WPK, channel between WAP gateway and Bank server is secured by using SSL which does not provide non-repudiation [2]. In NSI, authentication latency is 4.75sec which is not appropriate [9]. To overcome the limitations of NSI, PKASSO is proposed. But PKASSO is not feasible for mobile devices because of its complexity and implementation problems. SaPKI overloads computation intensive operations over GSM and CDMA networks. LPKI is designed for mobile devices which is counterpart of PKI used in desktop environment. Moreover, in LPKI, main focus is only on time spent while generation and verification of the signature. There is not any information about the complexity and computation intensive operations involved in certificate request creation process in PKCS#10 standard and authentication latency which are considered in our proposed design.

III. PROPOSED PROTOCOL

The purpose behind to propose this protocol is security, easy to implement and offloading the processor and memory intensive functions involve during certificate request generation, path verification and management.

A. Components of the System

The proposed security mechanism comprises of the following components. Functionality of each component is briefly described below.

a. Mobile Device

A mobile node requests for registration and verification to generate certificate.

b. Management Terminal
It is an interface use to register identities of the users manually by accessing web service of the registration server.
c. Registration/Verification Server
A server used for managing and verifying the identities of the users.
d. IDMS
A system used to manage identities and respective information of the clients.
e. Domain Level certificate Management Server
This is just like proxy server used to offload computation intensive functions from Client.
f. CA Server
A trusted certificate authority that is responsible to bind respective identities with their public keys to provide reliable authentications to its clients.

B. Processes of the Protocol

The proposed protocol is divided into three Processes, Registration, Verification and Certificate management, due to its complexity because it is being designed for light weight mobile devices. Abbreviation used in this section is presented in Appendix A (a).

a) Registration

IDMS, Registration/Verification Server and Registration Terminal are members of secure domain of an organization.

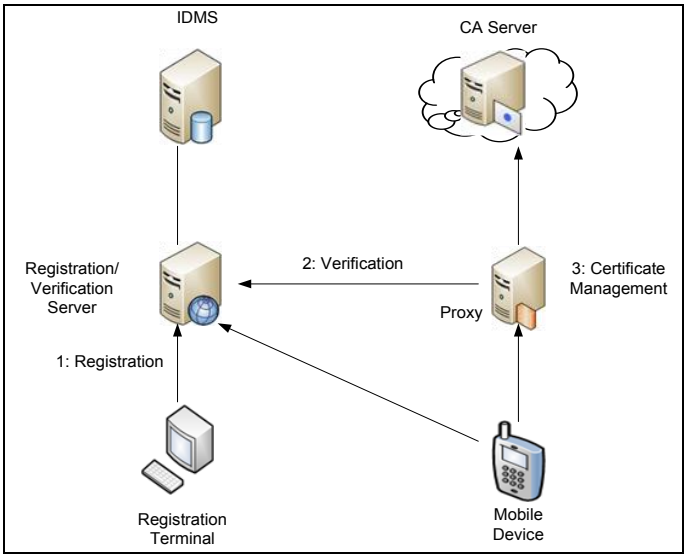


Figure 1: *Abstract Architecture*

There are two ways of registration. First is by visiting help desk management terminal. No user verification is required for registration as shown below Figure 2 (a). Second is by accessing web service by mobile device. In this case there will be policy to authenticate by its domain credentials as shown in Figure2 (b).
Standard SSL is implemented on the registration service for security.

User will login on secure web interface of the service and register its information and secure password. User can also register from mobile device by accessing secure web interface of the registration service.
M1: $Info_U|SP_U$
(Info:= Name, O: Organization, OU: Organizational unit, ID, Email, Country, City etc.)

M1 is the first message of the protocol, $Info_U$ is information of mobile user and SP_U is secure password.

Mobile user sends request to fetch registration data from Registration/Verification server (R/V Server) require for certificate. User hash of its ID, concatenate with nonce, encrypt with its secure password, concatenate encrypted message with plaintext ID and sends to the R/V server as shown by the message M2 in Figure 2 (c).
M2: ID_U | E [SP_U, (NO_U | H (ID_U)]

R/V server retrieve password from IDMS corresponding to the ID_U, decrypt the message. After that R/V server calculates the hash of ID_U and compares with received hash value. If both hash matches, integrity is ensured. After successful integrity confirmation, R/V Server creates distinguish name (DN_U), takes its hash, concatenate with received nonce and its ID_R, encrypt it with retrieved password SP_U, concatenate with plaintext DN_U and sends to User as shown by the message M3 in Figure 2 (c).
M3: DN_U | E [SP_U, (NO_U| ID_R | H (DN_U)]

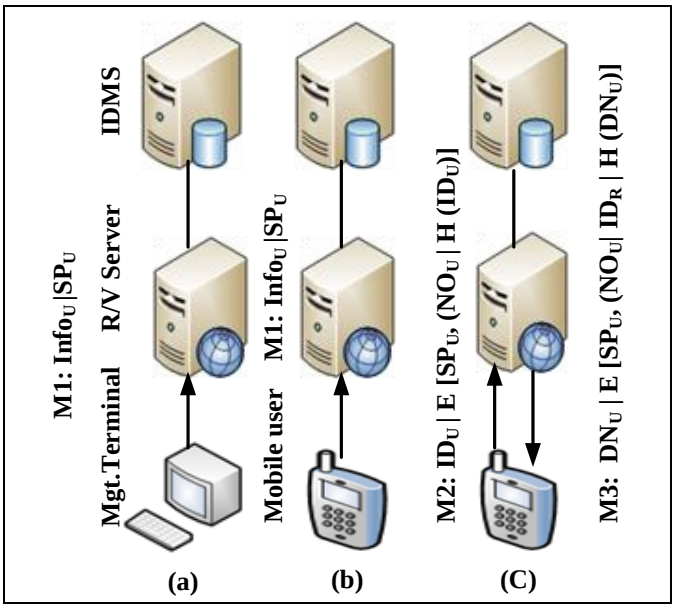


Figure 2: *Registration manual (a) By Mobile (b) Record Fetching process for Creating Cert. req. (c)*

User decrypts the message with its password, store received ID_R, compares nonce and hash with the calculated hash of DN_U to ensure integrity.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

b) Verification

If integrity is ensured, Client generate asymmetric key pair , takes hash of its public key and distinguish name, concatenate it with new nonce of User encrypts it with password and sends to Domain Level Certificate Management Server (Proxy) to create certificate request as shown by the message.

M4: (PU_U | DN_U) | E [SP_U, (NO_U | H (PU_U | DN_U)]

For verification, the Proxy Server forwards the same message to R/V Server.

M5: (PU_U | DN_U) | E [SP_U, (NO_U | H (PU_U | DN_U)]

R/V Server again retrieve the corresponding password from IDMS by using DN_U, decrypts the message, and compares the hash value with calculated hash of (PU_U| D_U) for ensuring the integrity of the message and authentication. If authentication is successful, R/V server sends "*Accept"* tag message along with encrypted message consist of received nonce and ID_R of Verification server otherwise "*Reject"* tag to Proxy Server.

M6: [Accept | E [SP_U, (NO_U|ID_R)]

After successful verification, Proxy Server generates "certificationRequestInfo value" (First part of PKCS#10: Contains encoded distinguish name of the subject, public key and other attributes), takes its hash, concatenate it with encrypted message received from Verification server and sends to Mobile Device for signing.

M7: [H (CRInfo.) | E [SP_U, (NO_U|ID_R)]

User (Mobile Device) decrypts the message, compares nonce with sent nonce NO_U and ID_R with stored identity in M3 of registration server. If verification is done successfully, User signs H (CRInfo.)) with its private key to avoid non-repudiation.

M8: E [PR_U, H (CRInfo.)]

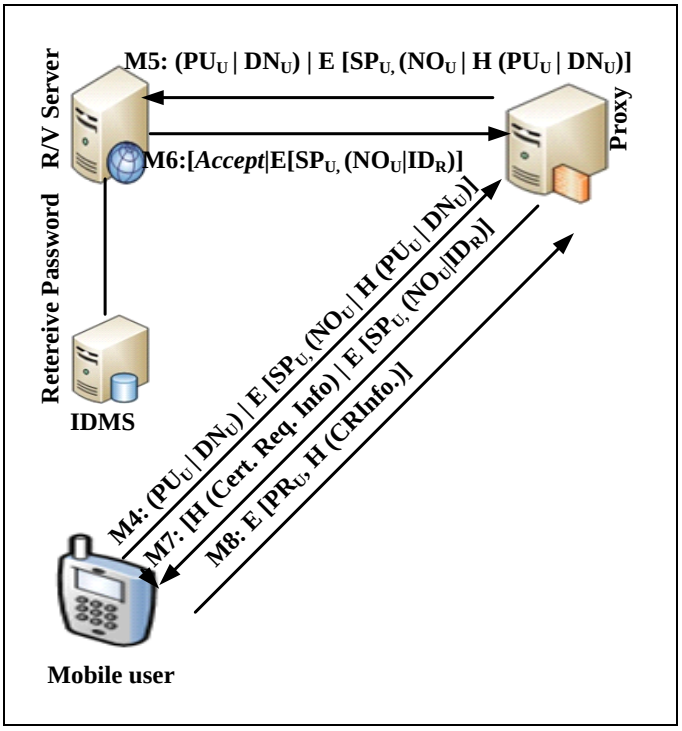


Figure 3: *Verification process*

c) Certificate Management

Proxy Server integrates the signature with CRInfo. (Certificate Request Info) and algorithm identifier to form certificate request in PKCS#10 standard and sends to CA Server for issuing certificate.

M9: [Signed PKCS#10]

CA server transforms the request into an X.509 public key certificate [10] by following the format standard PKCS#7 [11] and sends it to Proxy Server P.

M10: [Signed Digital Certificate by CA]

Proxy server sends that certificate to User.

M11: [Signed Digital Certificate by CA]

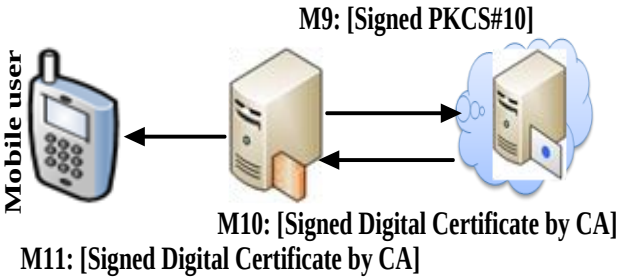


Figure 4: *Certificate Management Process*

IV. Formal Verification

Formal verification of the security protocols is very essential before their deployment because many protocols, considered secure but found insecure later like Needham-Schroeder Public Key Protocol [18].

In this paper, we are presenting Z based approach for formal verification and correctness of the protocol.

A. Z modeling of the protocol

Required notations are presented in Appendix A(b).

a) Dynamic Behavior as a set of Z Operations

User (U) -------> Proxy (P)

ΔInTransit
to = ⊥ ∧ to′ = P ∧ from′ = U
msg′ = (PU_U|DN_U) ⁀⟨enc [SP_U,(NO_U|H (PU_U|DN_U))]⟩

⊥ means it starts from no agent. msg′ is message after transition.

Proxy (P) -------- > Verifier (V)

ΔInTransit
to = V ∧ from′ = P
(∃ F: AGENT ; to′ = F ∧
msg′ = (PU_U|DN_U) ⁀⟨enc [SP_U,(NO_U|H (PU_U|DN_U))]⟩

Verifier (V) ---------- > Proxy (P)

ΔInTransit

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

to = V ∧ to´ = P ∧ froḿ = V
msǵ = (*Accept or Reject*) ⌢ ⟨enc [$SP_{U,}$ (NO_U|ID_R)]⟩

Proxy (P) --------- > User (U)
If verification is successful and accept message received to Proxy from Verifier, then

∆InTransit
to = U ∧ from´ = P
∃H:Hash•msǵ= H (CR info value) ⌢enc[$SP_{U,}$ (NO_U|ID_R)]

b) Formalizing the Attack

1. U Sends to F (Fake Agent)
2. F Sends to P
3. P proceed to V
4. V Reply to P
5. P Ack to F
6. F sends to U

-U Sends to F (Fake Agent)

∆InTransit
to = ⊥ ∧ to´ = F ∧ from´ = U
msǵ = (PU_U|DN_U) ⌢⟨enc[SP_U, (NO_U| H (PU_U|DN_U))]⟩

-F Sends to P
Fake user can intercept and modify the message. It can only tamper plaintext part (PU_U|DN_U) and can do nothing with encrypted part except create its own.
<u>If intruder is neither part of domain nor know any valid user ID</u>, attack cannot be successful. Lets intruder tamper DN_U like (PU_U|DN_U) with (PU_U|DN_F) or PU_U with PU_F or both like (PU_U|DN_U) with (PU_F|DN_F) Then

∆InTransit
to = F ∧ to´ = P ∧ from´ = F
msǵ = (PU_U|DN_F) ⌢ ⟨enc[SP_U, (NO_U|H (PU_U|DN_U))]⟩

- P proceed to V
Proxy server forwards the same message to verifier for verification.
ID_F ∉UID (Set of registered User IDs) and no password will be retrieved from IDMS.
ID_F is identity of intruder which is taken from DN_F (Distinguish name of intruder). So verification will be failed and Reject message will be sent to P and P will send this message to F.
<u>If Intruder is part of domain or knows any valid ID.</u>
In this case ID_F ∈ UID but corresponding password SP_F will remain unable to decrypt the message, so verification will be fail and attack will not be successful.
<u>If Intruder is part of domain or know any valid password and also create encrypted part of the message with its password Then,</u>

∆InTransit
to = F ∧ to´ = P ∧ from´ = F
msǵ = (PU_U | DN_F) ⌢⟨enc[SP_F, (NO_F|H (PU_U | DN_F))]⟩

In this case, IDF ∈ UID, corresponding password SP_F will successfully decrypt the message, integrity will also be confirmed, and so verification will be successful.
Dec [SP_F, (NO_F| H (PU_U | DN_F))]
= (NO_F| H (PU_U | DN_F))
= H (PU_U | DN_F) which is equal to the calculated hash of H (PU_U | DN_F)

- V Reply to P

∆InTransit
to = V ∧ to´ = P ∧ from´ = V
msǵ = (***Accept***) ⌢ ⟨enc [$SP_{F,}$ (NO_F |ID_R)]⟩

- P Reply to F

∆InTransit
to = P ∧ to´ = F ∧ from´ = P
msǵ = H(CR. Info.) ⌢ ⟨enc [$SP_{F,}$ (NO_F|ID_R)]⟩

- F Sends to U

∆InTransit
to = F ∧ to´ = U ∧ from´ = F
msǵ = H(CR. Info.) ⌢ ⟨enc [$SP_{F,}$ (NO_U | ID_R)]⟩

Intruder will send the message to honest user U for signing. User will be unable to decrypt the message. User will not sign the hash value, so attack will not be successful.

c) *Bypassing the Verifie***r**
<u>Intrude may deploy its own Proxy which bypass the verification process</u>
If Intruder intercepts the message and no matter modify the message or not and sends the hash of CR Info Value as message below.

∆InTransit
to = F ∧ to´ = U ∧ from´ = F
msǵ = H (CRInfo.) ⌢ ⟨enc[SP_U, (NO_U| H (PU_U | DN_U))]⟩

Intruder cannot send any message encrypted with its key because decryption will be fail that result in failure of the attack. In this case honest agent will successfully decrypt the message, receive same nonce but did not receive identity (ID_R) of Verifier that was received in fetching process. Finally this attack will also be fail.

V. RESULTS AND EVALUATION

After formal verification, protocol is implemented to obtain its results. Then these results have been compared with results of previous proposed solutions regarding Authentication latency, efficiency and operation time.

A. Specifications of implementing Environment

Host Machine

CPU	1.6 GHz (Dual CPU)	**RAM**	2.5GB

- Registration/Verification server is implemented on netbeans on host.
- My SQL server 5.5 for IDMS
- Mobile Client is implemented on Eclipse along with Android emulator 2.3

Virtual Machine

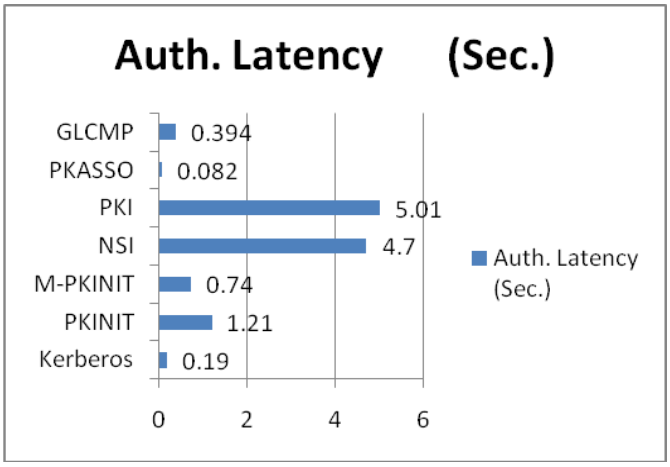
http://sites.google.com/site/ijcsis/
ISSN 1947-5500

CPU	Shared	RAM	512

- Proxy is implemented on Eclipse in virtual machine.

Protocols	Security Services			
	Authentication	Non-Repudiation	Digital Signature	Authentication Latency
Kerberos	Yes	No	No	0.19 Sec
PKINIT	Yes	No	No	1.21Sec
M-PKINIT	Yes	No	No	o.74 Sec
NSI	Yes	Yes	Yes	4.70 Sec
PKI	Yes	Yes	Yes	5.01 Sec
PKASSO	Yes	Yes	Yes	0.082 Sec
GLCMP	Yes	Yes	Yes	0.394Sec

Table 1: *Security Services and Authentication Latencies*

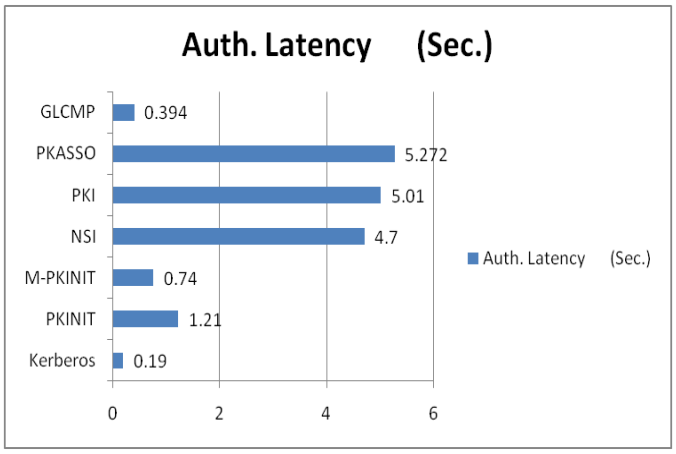


Authentication latency of proposed GLCMP is 0.394 sec which is 91%, 92% less than NSI and PKI respectively. As for as PKASSO is concerned, If we include 5.19 sec delegation time, our result is 93% efficient but if we do not include it then our authentication latency is 79% greater as shown in Table 3, but here we are also providing secure registration.

In PKASSO, 5.19 sec is consumed before delegation []. If we include this time then,

Protocols	Authentication Latency
PKASSO	5.272 Sec
GLCMP	0.394Sec

Table 2: *Authentication Latencies of the protocols*

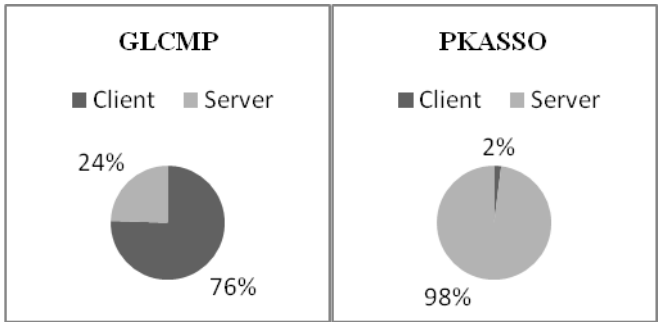


In our protocol, asymmetric key pair is generated on client side while in case of PKASSO; key pair is generated on delegation server.

Operation Time	KERBEROSE	M-PKINIT	NSI	PKASSO	GLCMP
On Client	0.024	0.518	4.72	0.066	2.78
On Server	0.036	0.333	0.51	3.253	0.9
Total	0.06	0.851	5.23	3.319	3.68

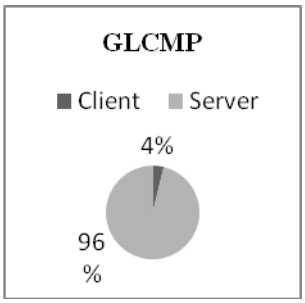
Table 3: *Operation time on client and server*

Here is comparison of PKASSO and GLCMP regarding operation time on client and server because PKASSO is the nearest competitor. In GLCMP, asymmetric key pair is generated on client so about 95% of the whole time is consumed in generating key pair. If we offload this task to proxy or any key management server, then whole picture will become as show in Fig 3 (c). Moreover, in GLCMP, secure registration is also provided.



(a) (c)

http://sites.google.com/site/ijcsis/
ISSN 1947-5500



(c)

Fig.3 *Operation time on client and server*

VI. CONCLUSION

Authentication, non-repudiation and digital signatures are very essential to be provided in mobile communication considering authentication latency, response time etc. GLCMP is very light weight; provide affordable authentication latency, based on generic security objects, provide secure registration of mobile users. It is also formally verified by using Z Notation modeling and concludes that it resists against man-in-the middle attack, replay attack, non-repudiation and provides certain level of security.

VII. REFERRENCES

[1] http://en.wikipedia.org/wiki/Public_key_infrastructure visited on April 2012.
[2] Fang LIU, QI YANG, "Study and Analysis of E-Commerce Security based on WPKI", DOI 10.1109/IITA. IEEE Workshop, 2008, 239.
[3] L. Zhu and B. Tung, RFC 4556: Public Key Cryptography for Initial Authentication Kerberos (PKINIT). IETF Network Working Group, 2006.
[4] Liang Cai, Xiaohu Yang, Chun Chen, "Design and Implementation of a Server-aided PKI Service (SaPKI)", Proceedings of the 19th International Conference on Advanced Information Networking and Applications (AINA'05) 2005 IEEE.
[5] RSA Laboratories. PKCS#10 v1.7: Certification Request Syntax Standard May 26, 2000.
[6] Zenel, B., A General Purpose Proxy Filtering Mechanism Applied to the Mobile Environment. Wireless Networks, 1999. 5: p. 391-409.
[7] S. Tuecke, V. Welch, D. Engert, L. Pearlman, and M. Thompson, RFC 3820: Internet X.509 Public Key Infrastructure (PKI) Proxy Certificate Profile. IETF Network Working Group, 2004.
[8] Mohsen Toorani, Ali Asghar Beheshti Shirazi, ―A Lightweight Public Key Infrastructure for the Mobile Environ, IEEE International Conference on Communication Systems (IEEEICCS'08), pp.162-166, Nov. 2008.
[9] Computationally Efficient PKI-Based Single Sign-On Protocol PKASSO for Mobile Devices Ki-Woong Park, Student Member , IEEE , Sang Seok Lim, Member , IEEE , and Kyu Ho Park, Member , IEEE
[10] ITU-T Recommendation X.509 (1997) | ISO/IEC 9594-8:1998, Information Technology- Open System Interconnection – The Directory: Authentication Framework
[11]RSA Laboratories. PKCS#7: Cryptographic Message Syntax Standards. Version 1.5, November 1993.
[12]M. Jalali-Shi and P . Ebinger , "Towards Efficient PKI for Restricted Mobile Devices," Proc. IASTED Int'l Conf. Comm. And Computer Networks, pp. 42-47, 2002.
[13] N. Modadugu, D. Boneh, and M. Kim, "Generating RSA Keys on a Handheld Using an Untrusted Server", Proceedings of the First International Conference in Cryptology in India, Lecture Notes in Computer Science, Vol. 1977, Springer-Verlag, Calcutta, India, 2000, pp. 271-282.
[14] N. Asokan, G. Tsudik, and M. Waidner, "Server-Supported Signatures", Proceedings of the Fourth European Symposium on Research in Computer Security (ESORICS), Lecture Notes in Computer Science, Vol. 1146, Springer-Verlag, Berlin, Germany, September 1996, pp. 131-143.
[15] J.-H. Han, Y.-J. Kim, S.-I. Jun, K.-I. Chung and C.-H. Seo, "Implementation of ECC/ECDSA cryptography algorithms based on Java card," Proceedings of 22nd IEEE International Conference on Distributed Computing Systems, pp.272-276, 2002.
[16] D. Hankerson, A. Menezes, and Scott Vanstone, "Guide to Elliptic Curve Cryptography," Springer-Verlag, New York, 2004.
[17] Benjamin W. Long, Colin J. Fidge, and Antonio Cerone, "A Z Based Approach to Verifying Security Proto cols", J.S. Dong and J. Wood cock (Eds.): ICFEM 2003, LNCS 2885, pp. 375–395, Springer-Verlag Berlin Heidelb erg 2003.
[18] H. Krawczyk, "HMQV: A high-performance secure Diffie-Hellman protocol (Extended Abstract)," Advances in Cryptology – CRYPTO'05, LNCS 3621, pp.546-566, Springer-Verlag, 2005.
[19] A. Harbitter and D.A. Menasce´ , "The Performance of Public Key-Enabled Kerberos Authentication in Mobile Computing Applica-tions," Proc. Eighth ACM Conf. Computer and Comm. Security, pp. 78-85, 2001.

VIII. APPENDIX A

(a)

U: Mobile user
$Info_U$: User Information
R/V Server: Registration/Verification Server
P: Domain Level Certificate Mgt. Server (Proxy)
C: CA Server
I: IDMS
PU_U: Public Key of Mobile user
PR_U: Private Key of Mobile user
SP_U: Secret Password of Mobile user
NO_U: User nonce
H: Hash Value
DN_U: Distinguish Name of Mobile user
E: Encryption
D: Decryption

(b)

(i) Set of Data Types

[AGENT] ::= U | P | V | ⊥ |
ITEM: =DN|Enc |Dec|H |Key |SP_U|NON|ID|CR Info
MSG== Seq ITEM

(ii) Subset of ITEMS:

NON: ℙITEM U: User P: Proxy
Key: ℙITEM V: Verifier
ID : ℙITEM ⊥: No Agent
Enc : ℙITEM CR Info: Certificate request info

Disjoint: ⟨DN, NON, Key, Enc, H, Dec, ID, CR Info ⟩
ITEM = DN **U** NON **U** Key **U** ID **U** Enc **U** H **U** Dec **U** CRInfo

(iii) Global State

InTransit
to : AGENT from : AGENT msg : MSG
Init
InTransit ---------- to = ⊥

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Efficient RSA Variant for Resource Constrained Environment

Seema Verma
Computer Science Department
Thapar University
Patiala, India .

Dr Deepak Garg
Computer Science Department
Thapar University
Patiala, India .

***Abstract*—The work in this paper is concerned with the memory consumption as well as the performance of RSA cryptosystem so that the most popular public key algorithm can be used efficiently in the resource constrained environment also. For this purpose, RSA variant, RC RSA, is proposed which results in low computational cost and low memory consumption. RC RSA is the improvement over dual RSA small e (based on less memory consumption). Mathematically, as compared to Dual RSA, RC RSA results in the increase of decryption speed by a factor of 9 and in implementation roughly by a factor of 6. On the other hand the encryption speed becomes as low as in standard RSA. Besides the computational speed up, RC RSA is proved to be more secure than the Dual RSA scheme.**

Keywords- cryptography; encryption; public key;security

I. INTRODUCTION

RSA [1] cryptosystem is widely used in many fields of communication. As symmetric key cryptographic algorithms (like AES [2]) are faster than public key cryptographic algorithms, but due to key exchange issues we cannot use symmetric key algorithms in any insecure network. So we cannot escape ourselves for using the public key cryptography. Public key cryptography is very costly in terms of computational and memory resources, still we have to use a public cryptosystem at least for transferring the shared key. There are many public key algorithms, but RSA [1] is the most popular one. RSA was proposed by Rivest, Shamir and Adleman in 1977. It is well known among cryptographers because of its simplicity, but this cannot be used efficiently in locations with constrained resources. The computational cost depends on the size of the encryption and decryption exponents. Standard RSA has less encryption cost and high decryption cost. Research has been done to improve the decryption side. The first improvement in decryption side can be found in [3] by Quisquater and Couverier which is known as RSA CRT (Chinese Remainder Theorem). Further improvements can be found in Batch RSA [4], MPrime RSA [5], MPower RSA [6], Rebalanced RSA [7] and RPrime RSA [8]. Comparison of the basic variants of is given in [9] by Boneh and Shacham(2002). Rebalanced RSA [7] and RPrime RSA[8] both increases the decryption speed by a good factor but at the cost of increased encryption speed. Both encryption and decryption exponents can't be optimized in one communication. In [10] and [11], the work is given to balance both encryption and decryption sides. The work to reduce the memory requirement by RSA is given in Twin RSA [12] and Dual RSA [13].

As RSA cryptosystem uses longer keys for the calculations, its memory consumption and calculation time is very large. These parameters make RSA unsuitable for resource constrained environment, like smart phones, banks with heavy load etc. Motivation of this paper is to reduce the memory requirement and improve the computational performance of the RSA cryptosystem, so that RSA can be efficiently used in a resource constrained environment.

The paper is designed as follows: in the second section basic RSA and its variants are described briefly. The third section contains the proposed scheme. Security analysis, complexity analysis, implementation details are described in fourth, fifth and sixth section

II. RSA AND ITS VARIANTS

A. Basic RSA

The RSA cryptosystem [1] is based on the multiplication of the two large primes. Its security depends upon the complexity of factoring a large composite integer. Two large random prime numbers are taken and multiplied for the generation of the key. But the reverse is not possible as factoring a large number is one way trapdoor function. Here the key generation, encryption and decryption algorithms are given for the understanding.

Key Generation Algorithm of RSA

1. Choose two large (say 512 bits) random prime numbers (p,q) and calculate $N = pq$ and $\phi(N) = (p-1)(q-1)$
2. Select an integer e, $(1 < e < \phi(N))$, such that $\gcd(e, \phi(N)) = 1$
3. Calculate the private exponent d such that $ed \equiv 1 (\mathrm{mod}\, \phi(N))$

Public key = (N, e) and Private key = (N,d),

The parameters $(p, q, \phi(N)$ are kept secret.

Encryption is done by calculating $C = M^e \bmod N$

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

For Decryption $M = C^d \bmod N$
Here C and M are the cipher text and plain text respectively.

As the decryption exponent d is usually very large, it makes the decryption very slow. Researchers have done a lot of work to improve the decryption side.

B. RSA with CRT

In RSA CRT [3], the decryption speed is enhanced by dividing the decryption work into two smaller parts. Then Chinese Remainder theorem (CRT) is used to merge the results.
Here in this variant, key generation & encryption algorithm remains the same as in basic RSA.

Decryption Algorithm

1. Two small decrypting exponents are calculated by $d_p = d \bmod p-1$ and $d_q = d \bmod q-1$
2. Two intermediate messages can be calculated by $M_p = C^{dp} \bmod p$ and $M_q = C^{dq} \bmod q$
3. The plaintext M can be calculated from M_p & M_q using Chinese Remainder Theorem(CRT)

Public key = (N,e) and private key = (N,d_p,d_q)
The decryption is done with the small parameters, d_p, d_q, p and q. Thus the operations involved are less time consuming as compared to the standard RSA.

C. MultiPrime RSA

MultiPrime RSA [5] proposed in 1998, is based on the composite number consisting of more than 2 prime factors.
Here the modulus N is calculated by k prime numbers $p_1, p_2 \ldots p_k$. The number of primes to be considered for the modulus depends upon difficulty of factoring the modulus N. Following is the method for the key generation:

Key Generation Algorithm

1. Select k random balanced prime numbers $p_1,\ldots,p_k$ such that $N=\Pi_{i=1}^{k} p_i$ is of required bit size.
2. Calculate e and d in the same way as in standard RSA with $\Phi(N)= \Pi_{i=1}^{k} (p_i-1)$
3. For k primes calculate the decrypting exponents, $d_i=d \bmod (p_i-1)$

Public key = (N,e) Private key = $(N,d_1,d_2...d_k)$
Encryption is done in the same way as in the standard RSA and decryption is done using CRT method for k primes ($p_1, p_2 \ldots p_k$) and k decrypting parameters ($d_1,d_2...d_k$). The performance of decryption method is better than RSA CRT because of the use smaller parameters as compared to RSA CRT.
.

D. Dual RSA

Dual RSA [13] proposed by Sun et.al, in 2007 is based on the generation of two RSA instances. The purpose of this variant is to reduce the memory consumption to store more than one RSA keys. It is used in the applications where two instances are required. This variant is used to generate two instances of RSA key pairs with same public and private exponents but different moduli. Two instances generated satisfy the following key equations:

$$ed = 1 + k_1 \phi(N_1)$$
$$ed = 1 + k_2 \phi(N_2)$$

Where k_1, k_2 are unique positive integers.

Dual RSA has three schemes in [13], namely Small-e, Small-d, and Generalized Dual RSA. The encryption and decryption methods are same for all these schemes. The encryption algorithm is same as standard RSA and decryption is done using the CRT method. Here we are concerned about only one scheme, i.e, Dual RSA Small e. The key generation method is given below:
Here, the number of bits of the public exponent $n_e < n/2$,

Key Generation Algorithm

1. Chose a n_e-bit random integer p_{11} and an $(n/2-n_e)$ bit integer p_{22}. If $p_1=p_{11}p_{22}+1$ is not prime, chose another set.
2. Chose an $(n/2-n_e)$-bit integer q_{22} and calculate $p_2=p_{11}q_{22}+1$, if p_2 is not prime repeat for another value.
3. Chose n_e-bit integer q_{11} such that $q_1=q_{11}q_{22}+1$ is prime.
4. Chose n_e bit integer e such that $\gcd(p_{11}p_{22}q_{11}q_{22},e)=1$ and calculate d and k_1 with the relation $ed=1+k_1(p_1-1)(q_1-1)$
5. Compute $q_2=k_1q_{22}+1$, if q_2 is not prime then repeat for another choice of e in step 4.

The modulus $N_1=p_1q_1$ and $N_2=p_2q_2$, $k_2=q_{11}$
Encryption is done in the same way as in the standard RSA and decryption is done in the same way as in RSA CRT.

III. Proposed Scheme

The motive of designing an RSA variant is to have less encryption, decryption and key generation time. Besides this computational cost memory is also a crucial factor. For achieving all these factors, MultiPrime feature is added to Dual RSA Small-e scheme, which increases the decryption and key generation speed. Due to the use of multiple primes in Dual RSA Small-e, security becomes more robust, so bit length of the public exponent decreases and ultimately encryption speed also enhances. Due to the use of Dual RSA, RC RSA consumes less memory where two instances of RSA are needed.
RC RSA is based on the two RSA key equations:

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

$ed=1+k_1*(p_1-1)(p_2-1)...(p_k-1)$ and
$ed=1+k_2*(q_1-1)(q_2-1)...(q_k-1)$
The method generates the two RSA instances such that both share the common public and private exponents. The parameters k_1, k_2, p_1, p_2,...p_k, q_1, q_2,...,q_k are calculated such as to satisfy these RSA key equation.

A. Key Generation Method (Generalized)

1. Public exponent e is selected as small random integer, say $n_e<n/3$. It might me sparse integer to have a less computational cost.
2. The prime factor p_1, p_2,...p_k, q_1, q_2,...q_k are computed as $p_1 = u_1 v_1 + 1$, $q_1 = u_1 v_2 + 1$, $p_2 = u_2 v_2 + 1$, $q_2 = u_2 v_3 + 1$,... $p_k = u_k v_k + 1$ such that u_1, u_1,...u_k with n_e bit and v_1, v_2,...v_k with (n/3-n_e)bit are random numbers satisfying $\gcd(e, u_1 v_1) = 1$, $\gcd(e, u_2 v_2) = 1$ ..., $\gcd(e, u_k v_k) = 1$
3. The parameters d and k_1 are computed such that $ed = 1 + k_1 (p_1 - 1)(p_2 - 1)...(p_k - 1)$.
4. Compute $q_k = k_1 v_1 + 1$, if q_k is not prime then repeat the process by taking another random choice for the calculation of the prime factor p_k.

Here N_1 is calculated as $N_1 = p_1 p_2 ... p_k$ and N_2 is calculated as $N_2 = q_1 q_2 ... q_k$, k_1 and k_2 are security parameters such that $k_2 = u_k$.

Mostly N is taken to be 1024 or 2048 bits, for this bit length modulus, k must be 3 [19, 20], here the algorithm is described for 3 prime factors.

B. Key Generation Method(k=3)

1. Public exponent e is selected as small random integer, say $n_e<n/3$. It might me sparse integer to have a less computational cost.
2. The prime factor p_1, q_1, r_1, p_2, q_2 are computed as $p_1 = u_1 u_2 + 1$, $p_2 = u_1 v_2 + 1$, $q_1 = v_1 v_2 + 1$, $q_2 = v_1 w_2 + 1$, $r_1 = w_1 w_2 + 1$ such that u_1, v_1, w_1 with n_e bit and u_2, v_2, w_2 with (n/3-n_e)bit are random numbers satisfying $\gcd(e, u_1 u_2) = 1$, $\gcd(e, v_1 v_2) = 1$ and $\gcd(e, w_1 w_2) = 1$
3. The parameters d and k_1 are computed such that $ed = 1 + k_1 (p_1 - 1)(q_1 - 1)(r_1 - 1)$.
4. Compute $r_2 = k_1 u_2 + 1$, if r_2 is not prime then repeat the process by taking another random choice for the calculation of the prime factor r_1.

here, $N_1 = p_1 q_1 r_1$ and $N_2 = p_2 q_2 r_2$, $k_2=w_1$

C. Correctness of the Key Generation Algorithm

As the RSA key equation is $ed = 1 + k\phi(N)$. The following way will prove the basic RSA key equation for RC RSA for k=3:

$$
\begin{aligned}
ed &= 1 + k_1 \phi(N_1) \\
&= 1 + k_1 (p_1 - 1)(q_1 - 1)(r_1 - 1) \\
&= 1 + k_1 (u_1 u_2)(v_1 v_2)(w_1 w_2) \\
&= 1 + w_1 (u_1 v_2)(v_1 w_2)(k_1 u_2) \\
&= 1 + k_2 (p_2 - 1)(q_2 - 1)(r_2 - 1) \\
&= 1 + k_2 \phi(N_2)
\end{aligned}
$$

Hence the keys generate valid key equations. In the same way the key generation method for generalized RC RSA can be proved.

D. Efficiency

The complexity of the new key generation method seems to be more complex because of the increase in the number of steps as compared to the Dual RSA Small-e scheme. The complexity is mainly concerned with the prime number generation. In the new scheme, short length primes are generated which takes less time as compared to the long length primes.

Encryption in RC RSA is done using small public exponent, which is the crucial factor in the encryption speed. Also, the value of the public key exponent e is chosen first. One can choose this value with low hamming weight (less no. of 1's). This will further reduce the encryption time because it reduces the computations involved in square and multiply method of the exponentiation.

Decryption speed is enhanced because of the operations involved with small prime numbers and small private exponents. Further details of the complexity analysis is given in section V.

IV. SECURITY ANALYSIS

In this section security of RC RSA is analyzed rigorously by considering many factors. Two important aspects are used in the cryptanalysis of RSA; first is Coppersmith lattice method and second is continued fraction method.

Lattice method: Coppersmith [15], Howgrave Graham [16], Jutla [17] and many other researchers have given their analysis of lattice attack. According to their method, for f(x,y,...z) be a linear polynomial with integer coefficients. For every ε>0, there is a positive integer M_0 and for every integer $M> M_0$, where M is relatively prime to at least one coefficient (not the constant), k linearly independent polynomials can be found and every root (x', y', ... z') of the given polynomial modulo M is also the root of the k polynomials modulo M, and if |x'|<X,|y'|<Y,.....,|z'|<Z and XY…Z<M, then (x', y',…z') is also a root of each of the k polynomials over the integers. The root (x', y',…z') can be computed if the k polynomials are

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

algebraically independent, which can be assumed in almost all the cases.

Continued fraction method: If x,y,z,w are positive integers such that $|x/y-z/w|<1/2w^2$ and gcd(x,y)=gcd(z,w)=1, then z/w can be calculated as one of the convergents of the continued fraction expansion of x/y.

Following are some of the attacks which can be questionable about the security of the current system. Each of the attack is proved to be ineffective in accordance with RC RSA.

A. Factorization Problem

Like MultiPrime RSA, RC RSA can be broken by factoring the modulus N. Once the private key d or multiple of Φ(N) is known, modulus can be factored by probabilistic method. According to [19, 20], the number of primes can be taken 3,3,4,5 as safe for modulus size 1024, 2048, 4096, 8192 bits respectively. It has been proved MultiPrime RSA for these factors is safe against ECM (for small factor) and NFS (for small modulus) factorization method. Hence in RC RSA, for modulus size $N = 1024$ and $N = 2048$, three balanced primes can be taken safely. No factorization algorithm works for this parameter. Security will be decreased if more number of primes are taken under these moduli sizes.

B. Finding k1 and k2(lattice method)

By using lattice method, the two security parameters k_1 and k_2 can be revealed if $n_e+n_d<7n/6+l/2$ or $n_e<n/6+l/2$, for $n_d\approx 1$.

Proof: RSA key equation

ed=1+kΦ(N) or

ed = 1+k(p-1)(q-1)(r-1)

=1+k{N-(pq-pr-qr+p+q+r-1)}

=1+k(N-t)

here t=(pq-pr-qr+p+q+r-1) with $n_t=2n/3$

The two key equations in RC RSA, $ed=1+k_1\Phi(N_1)$ and $ed=1+k_2\Phi(N_2)$, can be written as

$$ed = 1 + k_1(N_1 - t_1) \text{ and} \quad (1)$$

$$ed = 1 + k_2(N_2 - t_2) \quad (2)$$

Taking the difference of these two equations:

$$k_1(N_1 - t_1) = k_2(N_2 - t_2) \quad (3)$$

Few bits, l bits, of k_2 can be known by exhaustive search such that $k_2=k_l+k_m$, where k_l part is known and k_m is not known. Using this the equation (3) can be rewritten as:

$k_1N_1-k_mN_2+k_mt_2+k_lt_2-k_1t_1-k_lN_2=0$

$f(x,y,z)=N_1x-N_2y+z-C$

The roots of the polynomial are x', y', z'; where $x'=k_1$, $y'=k_m$, $z'=k_lt_2+k_mt_2-k_1t_1$, $C=k_lN_2$, C is a constant.

The polynomial can be solved for the upper bounds $X\geq x'$, $Y\geq y'$, $Z\geq z'$, provided that N is very large and XYZ < M, where

$M=\|f(xX,yY,zZ)\|_\infty = 2^{ne+nd}$

$X=2^{ne+nd-n}$

$Y=2^{ne+nd-n-l}$

$Z=2^{ne+nd-n+2n/3}$

Solving for XYZ<M

$n_e+n_d < 7n/6+l/2$ (4)

$n_e < n/6+l/2$, for $n_d\approx 1$ (5)

So for n=1024, l=80 and $n_d=1$; $n_e<211$ bits can reveal the two security parameters k_1 and k_2.

C. Finding k1 and k2 (continued fraction method)

By continued fraction method the two security parameters k_1 and k_2 can be computed if $n_e<n/6+l/2-\varepsilon$. Thus this method also give the approximate bound as discussed in the lattice method.

Proof: In RC RSA $N_1=p_1q_1r_1$ and $N_2=p_2q_2r_2$ can be rewritten as:

$$N_1 - 1 = u_1u_2v_1v_2k_2w_2 + u_1u_2k_2w_2 + v_1v_2k_2w_2 + u_1u_2v_1v_2 + k_2w_2 + u_1u_2 + v_1v_2$$

$$N_2 - 1 = u_1u_2v_1v_2k_1w_2 + u_1u_2k_1v_2 + v_1u_2k_1w_2 + u_1v_1v_2w_2 + k_1u_2 + u_1v_2 + v_1w_2$$

Dividing both these equations by $u_1u_2v_1v_2k_1w_2$:

$$\frac{N_1-1}{u_1u_2v_1v_2k_1w_2} = \frac{k_2}{k_1} + \frac{u_1u_2k_2w_2 + v_1v_2k_2w_2 + u_1u_2v_1v_2 + k_2w_2 + u_1u_2 + v_1v_2}{u_1u_2v_1v_2k_1w_2}$$

$$\frac{N_2-1}{u_1u_2v_1v_2k_1w_2} = 1 + \frac{u_1u_2k_1v_2 + v_1u_2k_1w_2 + u_1v_1v_2w_2 + k_1u_2 + u_1v_2 + v_1w_2}{u_1u_2v_1v_2k_1w_2}$$

Dividing these equations and performing a few steps of calculations, one can get:

$$\left|\frac{N_1-1}{N_2-1} - \frac{k_2}{k_1}\right| < \frac{2^{19}}{2^{n/3}}$$

According to the continued fraction method k_2/k_1 will be one of the convergent in the series of continued fraction expansion $(N_1-1)/(N_2-1)$, if

$$\left|\frac{N_1-1}{N_2-1} - \frac{k_2}{k_1}\right| < \frac{2^{19}}{2^{n/3}} < \frac{1}{2k_1^2}$$

$$k_1 < N^{1/6} + \sqrt{\frac{l}{2}} - \varepsilon$$

As $k < e$,

$n_e < \frac{n}{6} + l_m - \varepsilon$, this bound is approximately same as the bound received by lattice method.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

D. Small public exponent and known k_1 and k_2

In this method it is proved that even if k_1 and k_2 are known and the public exponent is small; exhaustive search attack will not work.

As $N_1 = p_1 q_1 r_1$ and $N_2 = p_2 q_2 r_2$, where

$p_1 = u_1 u_2 + 1$, $q_1 = v_1 v_2 + 1$, $r_1 = w_1 w_2 + 1$ and

$p_2 = u_1 v_2 + 1$, $q_2 = v_1 w_2 + 1$, $r_2 = k_1 u_2 + 1$

Writing and rearranging these equations,

$$N_1 - 1 = u_1 u_2 v_1 v_2 k_2 w_2 + u_1 u_2 k_2 w_2 + v_1 v_2 k_2 w_2 + u_1 u_2 v_1 v_2 + k_2 w_2 + u_1 u_2 + v_1 v_2$$

$$N_2 - 1 = u_1 u_2 v_1 v_2 k_1 w_2 + u_1 u_2 k_1 v_2 + v_1 u_2 k_1 w_2 + u_1 v_1 v_2 w_2 + k_1 u_2 + u_1 v_2 + v_1 w_2$$

In this case k_1 and k_2 are known and u_1, v_1, w_1 can be found by exhaustive search. The unknown parameters are u_2, v_2 and w_2, these three unknown parameters cannot be calculated using two equation only. Thus exhaustive search attack doesn't work for RC RSA.

E. Lattice based method with known k_1 and k_2

In this method using Lattice theory it is proved that known k_1 and k_2 do not reveal any information about the factorization of RC RSA modulus.

Considering again the equations:

$ed = 1 + k_1(N_1 - t_1)$ and

$ed = 1 + k_2(N_2 - t_2)$

Computing $k_1' = \frac{k_1}{\gcd(k_1, k_2)}$ and $k_2' = \frac{k_2}{\gcd(k_1, k_2)}$

such that $\gcd(k_1', k_2') = 1$

$$k_1'(N_1 - t_1) = k_2'(N_2 - t_2) \quad (6)$$

here k_1' and k_2' are known; only t_1 and t_2 are unknown.

Taking mod k_2':

$t_1 \equiv N_1 \bmod k_2'$

Assuming $C_1 = N_1 \bmod k_2'$, $t_1 = C_1 + \Psi k_2'$, here Ψ is unknown

$n_\Psi = \frac{2n}{3} - (n_e + n_d - n - \gamma)$, where γ is the bit length of the $\gcd(k_1, k_2)$.

Putting this value in equation (6):

$k_1'(N_1 - C_1 - \Psi k_2') = k_2'(N_2 - t_2)$

Taking this equation Modulo N_1:

$f_{N1}(x, y) = k_1' k_2' x - k_2' y + k_2' N_2 - k_1' C_1$

Here $x_0 = \Psi$, $y_0 = t_2$, the polynomial can be solved with lattice method, the roots can be found if $XY < N_1$;

$$\frac{2n}{3} - (n_e + n_d - n - \gamma) + \frac{2n}{3} < n$$

$$\frac{4n}{3} + \gamma < n_e + n_d$$

or

$$n_e > \frac{n}{3} + \gamma;$$

it means all the parameters can be known if the above inequality holds, for $n_e > n/3$, this is not the case in RC RSA. Public exponent is always taken less than n/3. Thus RC RSA is secure for low value of public exponent. .

F. Security Summary

It has been proved above that for $n_e < n/6 + l/2$ and $n_e < 211$, k_1 and k_2 can be computed by lattice method as well as continued fraction method. But it is also proved that with the computation of k_1 and k_2, no extra information is revealed, i.e., RC RSA can not be broken by exhaustive search or factorization of the modulus. The security of Dual RSA is analysed in [21, 22, 23], but none of the attack is effective in case of RC RSA (as discussed above). With all these discussion the public exponent can be taken as small as the public exponent in the standard RSA; i.e., $e = 2^{16} + 1$ without compromising the security of the system.

V. Complexity Analysis

In this section complexity of RSA variants are analysed in terms of performance and memory requirement. Current schemes are compared to RSA CRT and Dual RSA schemes. For the sake of simplicity N is considered 1024bits in this section. Table I is showing the comparison of the performance and memory consumption of the proposed schemes with other RSA schemes. The modular exponentiation [14] takes $(3/2)n_x * n^2$, n_x is the number bits of the exponent and n is the number of bits in the modulo. If the exponent is of type $2^{nx}+1$, then complexity comes out to be of special form $(n_x+1)n^2$.

TABLE I. Comparison of Performance Complexity of different RSA variants

	RSA Variants	Encryption Complexity	Decryption Complexity	Memory Consumed ($n+n_e+n_d$)
1	RSA	n=17 $17n^2$	n_d=1024 $1536n^2$	4n+34
2	RSA CRT	n_e=17 $17n^2$	n_d=1024 $384n^2$	4n+34
3	Dual RSA Small-e	n_e=296 $444n^2$	n_d=1024 $1536n^2$	3n+296
4	RC RSA	n_e=17 $17n^2$	n_d=1024 $170n^2$	3n+17

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

For basic RSA, e is usually taken to be $2^{16}+1$ and n_d with all probability comes out to be of n bits, so encryption complexity comes out to be $17n^2$ and decryption complexity comes out to be $3/2n^3$. In RC RSA, n_e=17, n_d=n. Here the decryption cost is calculated using CRT method. for the decryption in RSA CRT, the decryption complexity can be calculated by $2*(3/2)*(n/2)(n/2)^2$. In RC RSA, three secret parameters are used so decryption complexity by CRT can be calculated by $3*(3/2)*(n_d/3)*(n/3)^2$.

RSA cryptosystem uses three parameters which consume memory, public exponent n_e, private exponent n_d and the moduli n. If we have to store two instances of basic RSA all the parameters take their own memory space for each instance, i.e., 2n, $2n_e$ and $2n_d$ for moduli, public exponents and private exponents respectively.

The theoretical complexity comparison is shown in Table I. In the table the values are approximated with parameter n^2 (n=1024 where required) so that the variants can be better compared with one another. It is clearly shown in Table I that the decryption speed of RC RSA is theoretically enhanced by a factor of 9. If the decryption is done using CRT in Dual RSA Small e scheme, the speed of RC RSA is enhanced by a factor of 2.

The other major improvement is on encryption side. In the Dual RSA Small e scheme, due to security constraints the encryption complexity is very high. The length of the public exponent in Dual RSA Small-e is (n_e=296). In the case of RC RSA, after rigorous security analysis the encryption complexity comes out to be very small of the order of standard RSA. In RC RSA, the length of public exponent can be taken very small, say e=$2^{16}+1$.

In Table I, the encryption cost, decryption cost and memory consumption is shown to be better than Dual RSA scheme.

VI. IMPLEMENTATION

RC RSA is implemented on a laptop with 2.3GHz CPU and 5GB RAM. NTL [18] which is popular among researchers, with GMP using Cygwin tools on Windows 7 operating system is used. All the algorithms are run 100 times and the average is taken to record the results.

Table II shows the time taken by key generation, encryption and decryption algorithms by RC RSA for n=1024.

With moduli 1024 bits, the following values are recorded:

length(N1)=1024 bits
length(N2)=1024 bits
length(d)=1024 bits
length(e)=17 bits
e=65537
d=161461816393387084600238062954452367332075089168
849462538531109613337802394048843499736224632930914
069189618840110283677084021149990200033600751565713
434341010189801858250433031679281864771965094316674
585590071681302466917455817250334490127934483190807
072627251820428150774606537883231859537073500218624
3435855873
N1=124655110981215358630734637779732645342595008939
530760842377171458030811840292845354375329301766955
463191246363870869110880830955403288836924449063332
750939470367163535432915885995871896675093362573080
213233200787501457702506610051349730798675801567995
661016408955728637877220352935903670747845105529272
954801
N2=143213012410315739544253490848932773898903815489
103605814038995922603393724268880690263817585648458
685559292796774933892376672724293607510921489285049
842620378672749074047750322017353869464319193854084
896679045157035402447825161057765833762206160537682
784129663864770253056593254400067532523290981217324
610769
p1=287625285907859291358272990271535882498781349466
769332783385160892380056611626967176489200474870794
3553
q1=623525342037503755698147102846149691135189972927
542563826587371033238971246594720882002305146152496
9217
r1=695070596611088552344126767492229328714278151073
352995756321350385662842830477538342904596791351994
6801
p2=549377586639870437908642765725143047075397054826
022244913990752248958541919865710194157219854974749
7617
q2=434676671795355130876870333283899723681029674735
509565503517764804660115871057731469189720683006339
0657
r2=599715411724708978381403127647912279946262027980
531517364374226688725690570678759696294665896768682
5601
Time consumed by key generation algorithm=103.42 ms
Time consumed by encryption=0.172 ms
Time consumed by decryption=0.62 ms

The algorithm is also implemented with moduli 2048 bits, and the following values are recorded:

length(N1)=2048 bits
length(N2)=2048 bits
length(d)=2048 bits
length(e)=17 bits

TABLE II. COMPARISON OF THE RSA VARIANTS

RSA Variants here n=1024 k=3	Key Generation Time (ms)	Encryption Time (ms)	Decryption Time(ms)
RSA	100.8	0.134	3.2
RSA CRT	110.95	0.129	1.17
Dual RSA Small-e	315.9	1.09	3.9
RC RSA	103.42	0.172	0.62

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Time consumed by key generation algorithm=957.8 ms
Time consumed by encryption=0.31 ms
Time consumed by decryption=4.7 ms
Dual RSA Small-e and other RSA variants are also implemented using the same configuration for better comparative analysis. The results are shown in Table II.

VII. CONCLUSION

In several applications, like blind signatures and authenticity, there is the requirement of more than one RSA keys. Dual RSA solves this purpose. But Dual RSA cannot be applied directly in resource constrained locations. The work has been shown to get the improved variant of RSA that can be used for saving memory as well as computation cost. RC RSA is the improvement in the Dual RSA Small-e, besides the memory consumption it results in better key generation, encryption and decryption performance. As compared to Dual RSA, it increases the theoretical decryption speed by a factor of 9 and in implementation by a factor of 6. The encryption speed becomes as fast as in standard RSA. The scheme is proved to be secure against various attacks. RC RSA is useful in resource constrained locations, where memory and computation resources are less. This scheme is best suited to the locations like smart phones and banks with heavy load etc.

REFERENCES

[1] R. Rivest, A. Shamir and L. Adleman., "A method for obtaining digital signatures and public key cryptosystems", Communications of the ACM, 1978; 21(2):120-126.

[2] J. Daemen and V. Rijmen., "Rijndael, The advanced encryption standard", Dr. Dobb's Journal, 2001; 26(3):137-139.

[3] J J Quisquater and C Couvreur., "Fast decipherment algorithm for RSA public-key cryptosystem", Electronic Letters, 1982; 18:905–907.

[4] A. Fiat, "Batch RSA", Advances in Cryptology, 1989; 435:175–185.

[5] T. Collins, D. Hopkins, S. Langford and M. Sabin, "Public key cryptographic apparatus and method", US Patent #5, 1997; 848;159.

[6] T. Takagi, "Fast RSA-type cryptosystem modulo $p^k q$", Crypto'98, 1998;1462:318–326.

[7] M. Wiener, "Cryptanalysis of short RSA secret exponents", IEEE Transactions on Information Theory, 1990; 36(3):553–558.

[8] CAM Paixao, "An efficient variant of the RSA cryptosystem", 2003 preprints.

[9] D. Boneh, H. Shacham, "Fast variants of RSA", CryptoBytes, 2002; 5(1):1–19.

[10] H.M. Sun, M.J Hinek and Wu ME, "Trading decryption for speeding encryption in Rebalanced-RSA", Journal of Systems and Software, 2009; 82(9):1503-1512.

[11] S.D.Galbraith, C. Heneghan, J.F McKee, "Tunable balancing of RSA", In Proceedings of ACISP'05, 2005; 3574:280–292.

[12] A.K.Lenstra, BM De Weger, "Twin RSA", Progress in Cryptology-MyCrypt 2005; 3715:222–228.

[13] H.M Sun, M.E Wu, W.CTing and M.J Hinek, "Dual RSA and its security analysis", IEEE Transactions on Information Theory, 2007; 53(8):2922-2933.

[14] C. Vuillaume, "Efficiency comparison of several RSA variants", Master Thesis, Fachbereich Informatik der TUDarmstadt, 2003.

[15] D. Coppersmith, "Small solutions to polynomial equations and low exponent RSA vulnerabilities", Journal of Cryptology, 1997; 10:233–260.

[16] N. Howgrave-Graham, "Finding small roots of univariate modular equations revisited", In Proceedings of Cryptography and Coding, 1997; 1355:131–142.

[17] C.S Jutla, "On finding small solutions of modular multivariate polynomial equations", EuroCrypt, 1998; 158-170.

[18] V. Shoup, NTL: A Library for doing number theory. 2008, version 5.3.1. <http://shoup.net/ntl/>

[19] Compaq Computer Corporation: 'Cryptography Using Compaq MultiPrime Technology in a Parallel Processing Environment', 2000,Available online at ftp://ftp.compaq.com/pub/solutions/compaqmultiprimewp.pdf.

[20] M.J Hinek, "On the security of MultiPrime RSA", Journal of Mathematical Cryptology, 2008; 2(2):117-147.

[21] M.J Hinek, "On the security of some variants of RSA", Ph.D. thesis, University of Waterloo, 2007.

[22] S. Sarkar, S. Maitra, "Cryptanalysis of Dual CRT-RSA", IACR Cryptology eprint 2010

[23] S. Sarkar and S. Maitra, "Cryptanalytic results on Dual CRT-RSA and Common Prime RSA", Journal of Design and Codes Cryptography, 2013; 66:157-174.

AUTHORS PROFILE

Seema Verma received her B.Tech. degree in Computer Science and Engineering in 2001 and M.Tech degree in Computer Science and Engineering in 2007. She is currently pursuing her Ph.D degree in Computer Science and Engineering. Her research interests include information security and cryptography.

Deepak Garg has done his Ph.D. in the area of efficient algorithm design from Thapar University. He has more than 100 publications in International Journals and Conferences. He is working as Head , Computer Science department in Thapar University, India. He is chair, IEEE India Council Computer Society, India and Chair, IEEE India Council Education Society. He is chair, ACM SISACT, North India.

REAL TIME RECOMMENDER SYSTEM FOR MUSIC DATA

Mrs. Manjula Athani. *CSE,T.I.T,RGPV Bhopal , India*

Prof. Neelam Pathak *ITDept,T.I.TExcellence,RGPV,Bhopal, India*

Prof. Asif Ullah Khan *CSE,T.I.T,RGPV, Bhopal,India*

Dr.Bhupesh Gour *CSE,T.I.T,RGPV, Bhopal,India*

Abstract—Recommender system is able to identifying the n-number of users preferences and adaptively recommend music tracks according to user preferences. we are extracting unique feature tempo of each music using Marsyas Tool. Then we are applying BLX- α crossover to a extracted feature of each music track. User favorite and user profiles are included. This system have been emerging as a powerful technique of e-commerce. The majority of existing recommender systems uses an overall rating value on items for evaluating user's preference opinions. Because users might express their opinions based on some specific features of the item, recommender systems could produce recommendations that meet user needs. In this paper we presented a Real time recommender system for music data. Multiuser Real time recommender system combines the two methodologies, the content based filtering technique and the interactive genetic algorithm by providing optimized solution every time and which is based on user's preferences We can also share the favorite songs to other user hence it give better result and better user system.

Keywords-Recommender system, Interactive Genetic algorithm, Content Based filtering BLX- α

I. INTRODUCTION

The amount of information on ecommerce sites are increasing day by day. it becomes difficult for ecommerce users to choose the desired product from such an bulk of information. Recommender systems are an effective solution for it. Recommender System[1] are normally an information filtering technique that predicts the user items according to users personalized information obtained from results of algorithm. Incase of music website different categories of music may be available recommendation in this type of application will include recommending every user music according to the rating of the song and user profile and preferences.

II. SCOPE OF THE PROJECT:

The proposed system has great scope since this system can be used in almost ecommerce sites for music .we have chosen music since unlike other products one cannot just view and select the product. listening to all music may be tedious task.so this system can be of a great use in such cases. since this system is a dynamic one. Recommendation results for each user changes with the user preference.

3. PROPOSED WORK:

This system first extracts unique property of music tempo from the music file using a MYRSYAS TOOL This is Music Analysis and Retrieval Systems for Audio Signals. Marsyas uses ibt to find the value of tempo of each song. This extracted data is then stored on the database. Each stored property is analyzed using content based filtering[2],[3] and interactive genetic algorithm. The final step after applying genetic algorithm is displaying the items that are closest to the items which the user has given the highest rating. Using Euclidean distance formula the nearest possible music feature which are matching with the one generated by crossover step of genetic algorithm are matched and given as out for recommended items. Here a separate recommendation page is displayed where the top ten similar records matching which the two off springs generated is displayed.

Genetic algorithm procedure:

1. [Initialization] Randomly generate an initial population of solutions and evaluate the fitness function.
2. [New population] Create a new population by repeating the following steps .
   2.1[Selection] Select two parent solutions from a population according to their fitness (the better fitness, the greater the chance to be selected)
   2.2[Crossover] With a crossover probability cross over the parents to form a new offspring. If offspring is exact copy of parents then no crossover is performed.
   2.3[Mutation] With a mutation probability, At each position mutate new offspring.
   2.4[Acceptance] Place new offspring in a new population.
3. [Evaluation] Compute the fitness values for the new population of N solutions.
4. [Test] If the stopping criterion is met, stop, and return the best solution in current population.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

3.1 RELATED WORK

Recommender systems are internet-based software tools provides user with intelligent suggestions recommender systems for music data produce a list of recommendations. The main task of recommender system is how to recommend items tailored with user‘s preferences from the resources. According to the user favourite the recommender system provide the items corresponding with the user favourite. In order to resolve this matter there are two approaches in a recommendation system have been discussed in the literature i.e, content based filtering approach and the collaborative filtering approach.

In the content base filtering is based on the information and characteristics of the items that are going to be recommended. In this various candidate items are compared with items previously rated by the user and the best matching items are recommended.

However, the content-based system does not support the immediate changes in the potential interest of users. To eliminate these limitations, we combine the genetic algorithm approach and the content-based filtering in our proposed system .

Music Feature Extraction MARSYAS (Music Analysis Retrieval and Synthesis for Audio Signals) is a free software framework for audio analysis, synthesis and retrieval The main goal of Marsyas is to provide an extensible framework that can be used to quickly design and experiment with audio analysis and synthesis applications.

IBT [4] was developed in C++ and is freely available, under GPL licensing, in MARSYAS .ibt–standing for INESC-Porto Beat Tracker – is a tempo beat tracking system. Which will give the value of tempo[5]. The slow tempos are at the rate of 63,72,and 80 these tempo are most effective on sad songs. The fast tempos very strongly affect the happy-gay and vigorous groups and at the rate of 102,104,112,152.

3.1.1 Interactive Genetic Algorithm:

A Genetic algorithm is a search technique used in computing to find true or approximate solution to optimization and search problems. Genetic algorithms[6],[7] belong to larger class of evolutionary algorithms which generates solutions to optimization problems[8] algorithm use technique inspired by natural evolution, as inheritance, mutation, selection and crossover.

4.SYSTEM OVERVIEW

The recommender system described in this paper is based on the genetic algorithms. The content-based filtering technique is applied to generate the initial population of genetic algorithm. In the proposed system, we employ the interactive genetic algorithm so that the users can directly evaluate fitness value of candidate solution themselves. After the evaluation , our system can recognize and recommend items tailored with different user preferences.

The recommender system is divided into three phases: feature extraction phase, evolution phase, and interactive Genetic algorithm phase. The MYARSYAS software is provided with music file which extracts unique property of music tempo. This extracted data is then stored on the database. Using the content based filtering and interactive genetic algorithm the stored data is analyzed. After analyzing records, the system recommends items appropriate to users own favourite.

The user is get general list from which users can select the audio tracks, listen to it and give rating list and user favourite list where that user has given highest rating to the audio tracks.

4.1 Phases of Genetic Algorithm

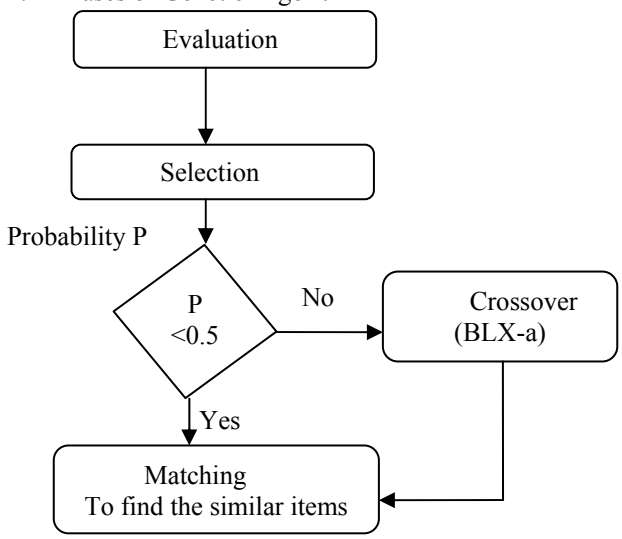


Fig 1 -The process of Interactive GA phase.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

4.1.1 The following are phases of generic algorithm are as follows:

4.1.1.1 Selection phase –Using MYRSYAS software Music features are extracted . In this system Truncation selection [9] is used, Those records which fall below threshold value are not selected and are ignored. The selected ones form the initial population for the genetic algorithm, where these records value are used in the next phase of this application.

4.1.1.2 Crossover phase :TheBLX–α crossover algorithm is used since extracted features are real numbers. Hence crossover is performed with this algorithm resulting in new generation.

Blend Crossover (BLX-α) was proposed by Eshelman and Schaffer(1993).It is reported that BLX (α=0.5) performs better than BLX operators with other α value.This algorithm is used to generate new off springs after the crossover step.

Crossover Algorithm: BLX- α

1. Select two parents X(t) and Y(t) from a parent pool
2. Create two offspring X(t+1) and Y(t+1) as follows:
3. fori = 1 to n do
4. di=|xi(t)-yi(t)|
5. Choose a uniform random real number u from interval <min(xi(t),yi(t))-α*di, max(xi(t),yi(t))+ α*di>
6. xi(t+1)=u
7. Choose a uniform random real number u from interval <min(xi(t),yi(t))- α*di, max(xi(t),yi(t))+ α*di>
8. yi(t+1)=u
9. end do

4.1.1.3 Matching phase- This phase finds the similarity between music features stored in database to the newly generated music features. The system recommended itemswhich are similar .
This phase uses Euclidean distance between two offspring and distance between each feature of the two offspring is calculated, resulting value is used to match the records stored in the database. Those records are compared with the resulting value which have highest rating given by the user.
Euclidean Formula:

$$d[] = \sqrt{\sum_{k=1}^{n} (x_{ik} - x_{jk})^2}$$

Where iand j are two items and k is the length of each music property. n is number of property.

5.THE EXPERIMENT

In this we describe the implementation of our proposed system and experiment results of n-users can dynamically register and give ratings.

5.1 Proposed Implementation:

We incorporate with this system, which is implemented in .NET the information gathered from Feature Extraction Phase. We then construct a website providing an experimental environment to make it easy for the user .
The website provides essential information such as artist name, songs title category, user count, give rating and overall rating; user favorite and user profiles[10],[11],[12], are included. users can rate their preferences about each music item by clicking the corresponding icon. Each time a user evaluates a page of n-items[13][14]. On any page any user can rate it and overall rating we get it. The initial page is statically generates according to database. The successive page is constructed based on the user evaluation. Dynamically n- number of songs can be added. One user can share their favourite to the other user. Below table shows the experiment result as shown below diagram.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

5. RESULT OF THE PRAPOSED WORK:

Song id	Artist	Title	Category	User Count	Tempo	Music
11	Sonu Nigam	Ai Zindagi	Sad	2	82.0000	Listen
23	Prakash	Govinda re	Dance	4	85.0000	Listen
35	Akon	dance floor	Pop/Remix	3	85.0000	Listen
10	Blaaze	Hosanna	Sad	9	86.0000	Listen
30	Neeti Mohan	Darbadar	Dance	2	86.0000	Listen
34	Sonu Nigam	Ramaiya	Romantic	4	91.0000	Listen
24	Avdut Gupta	Hey Lomboder	Happy/Fun	5	92.0000	Listen
33	Sayira	Ektha Tiger	Pop/Remix	3	92.0000	Listen
36	Sherya	Yarayara	Romantic	4	92.0000	Listen
6	Neha Basin	Dhunki	Dance	7	96.0000	Listen

Fig 2- Result of Experiment.

6. CONCLUSION

In this paper we presented a real time recommender system for music data. In this system is able to identifying the n-number of users preferences and adaptively recommend music tracks according to user preferences by applying BLX α crossover to extracting features of each music track. Thus we incorporated the main Interactive genetic algorithm based engine with content based filtering method.In this system User favorite and user profiles are included. On any page user can rate and overall rating we get it. We can also share the favorite songs to other user hence it give better result and better user system. According to subjective decision This system enables n-user can register and give ratings hence it give better result and better user system.

REFERENCES:

[1]Hung-Chen,ArbeeChen,"*A Music Recommendation System Based on Music and User Grouping*" 2005 Springer scienceJournal of Intelligent Information System 24:113-132.

[2]Byeong Man Kim & Qing Li. Chang Seok Park &Si Gwan Kim &JuYeonKim"*A New Approach for Combining Content-based and Collaborative filters*": J IntellInfSyst (2006) 27: 79–91

[3]Jun Wang1, Arjen P. de Vries1,2, Marcel J.T. Reinders1 "*Unifying User-based and Item-based Collaborative Filtering Approaches by Similarity Fusion*" (2006).

[4] Joao Lobatooliveira Fabien Gouyon"*IBT: Areal-Time Tempo and Beat Tracking System*".International Society for Music Information Retrieval (2010)

[5]Kate Hevner"*The Affective Value of Pitch and Tempo in Music*"The American Journal of Psychology,vol-49,(May-2013).

[6]Chein-Shung Hwang "*Genetic Algorithms for Feature Weighting in Multi-criteria Recommender Systems*" Journal of Convergence Information Technology Volume 5, Number 8, October 2010,PP 126

[7] Hyan-Tae Kim Eungyeongkim,"*A Recommender System Based on Genetic Algorithm for Music Data*",Intenational Conference on Computer Engineering and Technology. vol-6,2010,PP 415.

[8] Leticia C. Cagnina and Susana C. Esquivel" *Solving Engineering Optimization Problems with the Simple Constrained Particle Swarm Optimizer*"Informatica 32(2008) 319–326

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[9]Rakesh Kumar, Senior Member, IACSITandJyotishree, Member, " *Blending Roulette Wheel Selection & Rank Selection in Genetic Algorithms*" IACSITInternational Journal of Machine Learning and Computing, Vol. 2, No. 4, August 2012

[10]SachinBojewar and Jaya Fulekar "*Application of Genetic Algorithm for audio search with Recommender System*" International Journal of Advanced Computer and Mathematical Science . ISSN 2230-9624.Vol 3, Issue 2, 2012, pp 224-226

[11] MarcosA.Domingues*"Combining usage and content in an onlilne Music Recommender System for Music in the Long Tail*"April16-20,2012 ,Lyon,France.PP 927.

[12] NamdoBadheDivya Mishra, Chancha l Joshi, NehaShukla" *Recommender system for music data using genetic algorithm*" International Journal of Innovations & Advance me nt in Computer Science IJIACS ISSN 23478616 Volume 3, Issue 2April 2014

[13]ManjulaAthani, NeelamPathak, AsifUlhaKhan "*Recommender System Based on Genetic algorithm for songs on Web*"International Journal of Advanced Research in Computer Science and Software Engineering.ISSN:2277128X vol-3,Issue 12,Dec-2013.

[14] ManjulaAthani, NeelamPathak,AsifUlha Khan "*Dynamic music Recommender System Using Genetic algorithm*"International Journal of Engineering and Advanced Technology ISSN:2249-8958,vol-3,issue-4,April-2014.

ICT AS A TOOL FOR IMPROVING INFORMATION FLOW AMONG LIVESTOCK STAKEHOLDERS. A CASE STUDY OF TANZANIA.

Gladness Mwanga George
Nelson Mandela African Institution of Science and Technology

Fatma Simba,
University of Dar es Salaam,

Zaipuna O.Yonah,
Nelson Mandela African Institution of Science and Technology.

***Abstract*- Traditionally, extension services are used as a means of conveying to rural areas knowledge derived from different studies and surveys. These extension services are an important way to help livestock farmers to improve their lives and production methods, which in turn leads to an increased and improved livestock and livestock products. Nowadays, extension services deliver far more beyond the traditional role. They involve helping farmers to form collaborative groups, solving marketing challenges, etc. This fact has been confirmed by the study reported in this paper. The main objective of the study was to evaluate the current recording system and information flow among livestock stakeholders, and how ICT it has been used as a tool to bridge the gap of information deficit.**

This paper presents an analysis of data collected from 15 wards from Arumeru district, Arusha region – Northern Tanzania including data from; district council, livestock farmers, extension officers and researchers from three institutions, Livestock Training Agency (LITA), National Artificial Insemination Centre (NAIC) and Nelson Mandela African Institution of science and Technology (NM-AIST).

The results reveal that the current recording system is poor and there is a gap in the flow of information among the stakeholders of livestock farming. This gap has significantly contributed to the deterioration of access to extension services to livestock farmers. Along with other factors, this gap is also attributed to the researchers, who publish their research findings through various online journals. The gap signifies that it has been hard for extension officers to fulfil their roles due to a large number of farmers they serve and the challenge of traveling long distances to deliver such services to farmers. Based on the results of this analysis, the paper concludes with a set of identified approaches on how ICT can play a part in minimizing the gap that has been found for increasing the efficiency of the extension services in reaching livestock farmers.

Key words: ***ICT, information, Extension services, livestock stakeholders, record keeping.***

1. INTRODUCTION

Information and communication Technologies (ICT) are considered to bring economic and social development, with the benefits of reaching even those who do not themselves have first-hand access to them[1].ICT can enhance knowledge sharing and improve access to information [2]. While in agriculture, extension officers are responsible in conveying knowledge and scientific findings to rural areas for the purpose of improving the lives of rural livestock keepers [3]. In other term are refereed as intermediate channel between a farmer and a researcher. Apart from that, they also involve in animal husbandry, help farmers to create working groups and cope with other challenges such as marketing [4].

In delivering extension services (animal husbandry), extension officers also need information from farmers, such as animal profile information to give informed advice, hence farmers' recording keeping is crucial aspect in delivering extension services. Record keeping for livestock is a task of collecting, maintaining, and utilizing collected records [5]. Collected data from farmers is used as a management tool to undertake extension services, performance evaluation, keep proper health records, accurately measure production and reproduction, and perform other important management functions required to run an effective and efficient farm enterprise [6].In capturing data there are number of steps including recording on the form or computability devices, scanning the written document. And data to be stored can be in written form, image, videos or audio [7]. But in keeping records it is very important to consider how data will be extracted later on. Computer is among the tool which facilitate easily storage and retrieve of information compare to paper based [8].

In Tanzania, veterinary support is provided mainly through government agents at village, ward and district levels [9]. However, this sector is challenged with various problems, including, inadequate cooperation between stakeholders, insufficient expertise and weak researcher/extension officer/farmer linkage (National Livestock Policy of 2006). And often veterinary extension officers lack the skills and means to improve livestock keeping

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

and most of them are not specialized in livestock production[9].

This paper report about a study conducted for the purpose of establishing and identifying challenges facing the current recording system and flow of information among livestock stakeholders, and how ICTs have tried to bridge the gap. Data was collected through interviews from different stakeholders and from site visits to three livestock institutions. Data was also collected through questionnaires which were distributed to farmers and extension officers in Arumeru District in Arusha Tanzania. Also, during the survey qualified information was gathered through observation. Questionnaires targeted at obtaining insight knowledge on how data is stored and how information flows among different stakeholders (livestock farmers, extension officers, marketing and researchers etc.) and challenges which exist in current recording system and in connection with the flow of information among them.

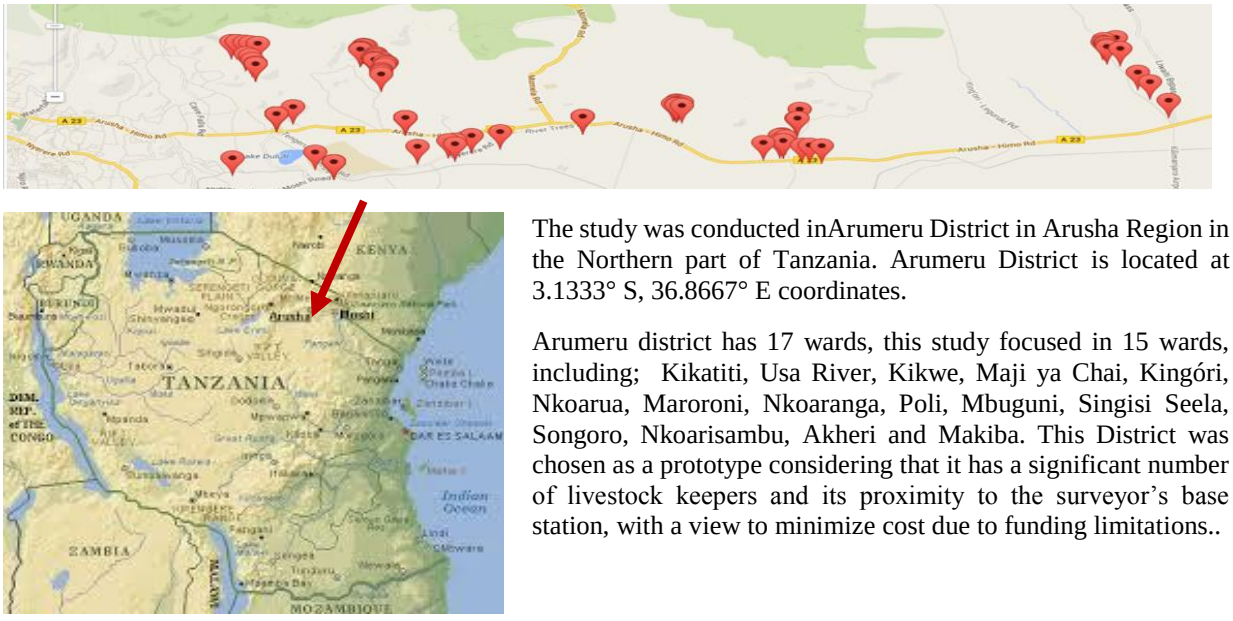
Through literature review it was identified that ICTs are, believed to reach the unreached [10] and information is a crucial factor in modern farm management [12] . A number of projects have been implemented in Tanzania and across the world to offer solutions that improve the ways through which farmers are reached and have access to information. These projects include those which provide information on marketing and best practice of farming through various media such as radios, computers and mobile phones.

In this paper we begin with an overview of the information flow and record keeping in a livestock sector, existing challenges and how ICT is proven to be a linkage opportunity to bridge identified gap. **Section two** of the paper provides an overview of the methodology used including; the study area and research design. **Section three** presents the obtained results and discussions. **Section four** provides an overview on how ICT is used to bridge the identified gap and **section five** contains the conclusion.

2. METHODOLOGY

This section present the methodology used to conduct the study reported in this paper. Including description of the study area, research design sample size and sampling techniques.

2.1: Description of the study area



The study was conducted inArumeru District in Arusha Region in the Northern part of Tanzania. Arumeru District is located at 3.1333° S, 36.8667° E coordinates.

Arumeru district has 17 wards, this study focused in 15 wards, including; Kikatiti, Usa River, Kikwe, Maji ya Chai, Kingóri, Nkoarua, Maroroni, Nkoaranga, Poli, Mbuguni, Singisi Seela, Songoro, Nkoarisambu, Akheri and Makiba. This District was chosen as a prototype considering that it has a significant number of livestock keepers and its proximity to the surveyor's base station, with a view to minimize cost due to funding limitations..

Figure 1: Description of the study area

2.2: Research design

The use of questionnaires, interviews, site visits and observations provided the grounds for obtaining qualified information. Questionnaires were distributed to livestock keepers and extension officers in 15 wards of Arumeru District. The questionnaires had different questions for livestock keepers and extension officers. The site visits to two livestock research Institutions in Arumeru district provided a clear picture on how livestock researchers perform their duties, and the interviews which were conducted to extension officers and the researchers from the Research Institutions, provided us with information on how they cooperate.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

2.3: Sample size and sampling techniques

The Arumeru District has 102,134 livestock farmers (according to 2012 National census) [11], we have use random sampling where we selected 210 livestock keeper to represent the rest, 15 extension officers one from each ward because one ward have two to three extension officers and 6 livestock researchers. Researchers were selected from three Research Institutions. The selected Institutions were:

1. Nelson Mandela African Institution of Science and Technology (NM-AIST)
2. Livestock Training Agency (LITA) and
3. National Artificial Insemination Centre (NAIC).

LITA and NAIC are Institutions which are specialized in livestock activities and are centre for training extension officers, while NM-AIST is a research Institution whose activities, amongst others, include livestock researches. Questionnaires with different questions were given to 210 livestock keepers and 15 extension officers. All questionnaires were successfully filled in by the use of the Open Data Kit (ODK) as well as printed questionnaires.

In addition, face to face interviews were conducted on researchers' policy makers and some extension officers. The interviews aimed at getting additional information that could elaborate more on what was obtained from the questionnaires. Also, site visits were made to two livestock research Institutes (LITA) and (NAIC). The visits were intended to observe the presence of resources for collecting and sharing data among livestock stakeholders. The collected data were organized and analysed by using the Statistical Package for Social Sciences (SPSS) and Open Data Kit (ODK).

3. RESULTS AND DISCUSSION

This section present the result obtained and the discussion. From the survey it is reported that, farmers still depend on extension officers for information concerning their livestock welfare. But while this stands to be true, extension officers, on the other hand, are not able to serve effectively all livestock keepers due to various challenges that they are facing. These include; serving a large number of livestock keepers who are located far apart over a wide area thus demanding them (extension officers) to walk long distances, which militate against serving all Livestock keepers accurately.

Consequently, some of the livestock keepers do not receive updated information and research findings frequently and on a regular basis. There is also the problem of getting information and research findings from Researchers after they have completed their studies and research. Most of them publish their work in online journals and other publications sites which are often not easily accessible by many of the stakeholders, and even for those who are capable of using modern communication facilities, find it difficult to retrieve the information due to lack of proper and efficient communication infrastructure. Also the current recording system is paper based which is associated with major challenges such as memory loss face difficulties in perform data analysis. As a result livestock keepers need a good system for collecting and analysing their daily farm data and information flow need to be restructured.

3.1: Type of information shared among livestock stakeholders

It is extremely important for the livestock keepers to have access to current livestock information, because such information helps them to make informed decisions concerning livestock keeping practices, marketing etc. [12]. The results obtained show that 31.4% of livestock keepers demand access to livestock information on different aspects of modern livestock keeping practices, disease outbreak, pharmaceuticals, feed and fodders, marketing insemination and vaccination to mention a few. The results were supported by Extension Officers, who, when asked what type of information is frequently requested by livestock keepers, they confirmed the above mentioned results.

Similar results are reported in [9], that knowledge and information on the best pastures, crossbreeding, animal husbandry and food is rarely available to livestock farmers and/or producers and , there are limitations related to access and availability of information relating to livestock diseases.

3.2: How do livestock stakeholders obtain information and how it is shared between them (Extension Approach)

Figure 2 and Figure 3 show results of the survey on how livestock stakeholders obtain information and how they share the same. It is recorded that the majority of livestock keepers (67.6%), obtain livestock education from extension officers while a few of them get the information from radios, televisions and printed media, and yet others get it through talking with their neighbours. Extension officers, on the other hand, acquire their knowledge from training they get from livestock training Institutions as well as from brush-up seminars, which are conducted occasionally. The results show that 63.3% of extension officers depend on seminars/training as one of the ways they use to gain knowledge. They also acquire knowledge through the use of internet and by having person to person communication with other extension officers via their mobile phones.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

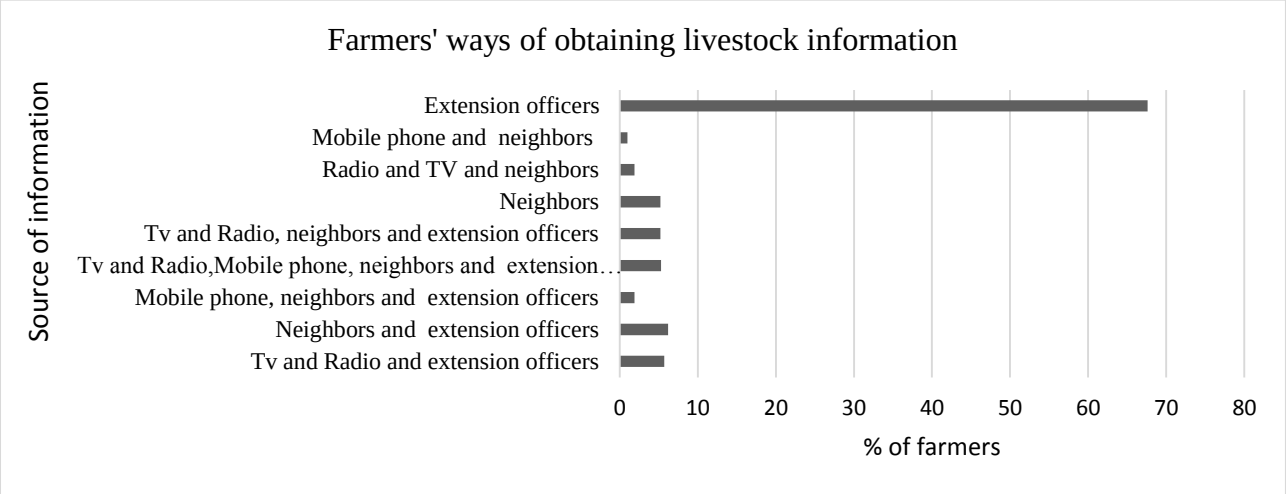


Figure 2: Farmers' ways of obtaining livestock information

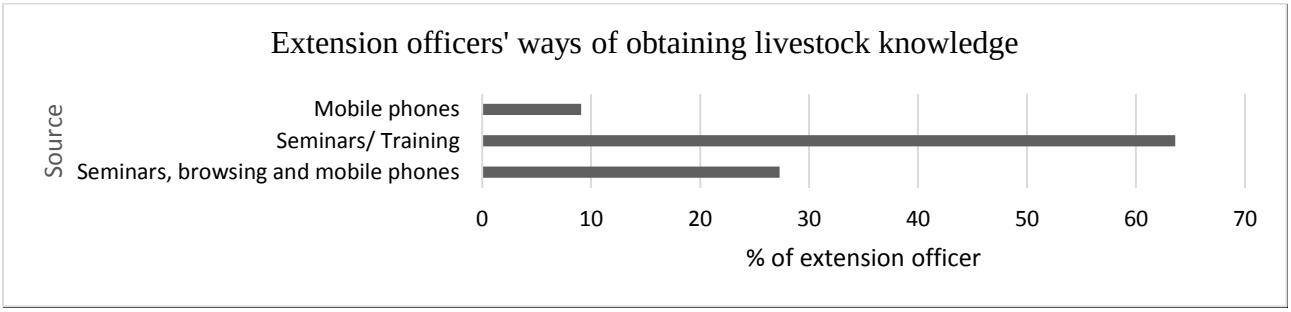


Figure 3: Extension officers' ways of obtaining livestock information

Table 1: Number of farmers that one extension officer serves

% Extension officers(respondents)	9.1	9.1	9.1	9.1	9.1	18.2	18.2	9.1	9.1
Number of Farmers	170	376	378	436	600	800	2800	4960	8729

However, the majority of the livestock keepers who responded to the questionnaires confirmed that they do communicate with extension officers. But, from the survey, it has been observed that more than 52.4% of them do communicate only when they have problems. This style of communication has a negative impact on the recommended way of keeping livestock because farmers need to be updated as frequently as possible on issues that have direct impact on productivity e.g. how to feed animals, changes in weather conditions, quick notification in case of disease outbreak, etc.

But extension officers, on the other hand, stated that even though livestock keepers report to them, on occasions when they have problems with their livestock, they would not be able to get the best and most effective service because of the problems and challenges that extension officers are facing. As already mentioned earlier, extension officers are overloaded by a very large number of clients who are scattered across a large area as in Table 1 making it difficult for them to make periodic and regular visits to each livestock keeper. This problem coupled with the absence of efficient and reliable transport facilities that makes it difficult for extension officers to provide effective, reliable and timely services to livestock keepers.

The government is quite aware of the existing problems and challenges which affect extension services, however, due to budgetary constraints, it has not been able to allocate enough funds to support extension services. The problems and challenges

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

that exist now are bound to persist if no efforts are done to alleviate them and extension officers will continue to face the problem of not being able to reach and provide proper and efficient services to large numbers of pastoralists [9]. Also, serving farmers at village level is a challenging task that requires not only a sufficient number of committed extension officers, but also facilitation in terms of creating a conducive working environment and support [13].

However, different researches are being conducted to address various livestock issues, but from this study it has been revealed that most of the research findings are published in different journal/websites where they cannot be accessed by other stakeholders due to poor communication infrastructure. Posted information also is not in simplified form for livestock keepers to understand.

Not only that, but also marketing information is very important for any type of business. The livestock keepers, likewise, needs, to have marketing information relating to the markets into which they sell their products. The current situation is that, livestock keepers have very poor access to marketing information. As such, they have little knowledge of the market situation in terms of, prevailing market prices, potential customers, transport costs, and auction dates/places. In the absence of marketing information, livestock keepers will have to rely on the unfavourable prices and other terms of business that are offered by the buyers. The current mode of communication/contact between buyers and sellers is through direct (face to face) encounters and by the use of mobile phones.

The linkage between farmers and buyers, the result of this survey shows that 40% of livestock keepers make direct (face to face) contact with buyers, while 32% of them contact buyers by mobile phones whenever they want to sell their animals or animal products. One of the study conducted in Tanzania [14], identified that marketing information is useful for livestock farmers whereby it provides more potential to identify market opportunities. However, the main government websites www.livestockinfo.go.tz, www.limstz.net, that were created for the purpose of providing livestock and agriculture information to the stakeholders, have failed to deliver fruitful results and the websites are currently not actively operational, [9]. This directly affects livestock keepers who still need marketing information.

In accessing marketing information respondents (farmers) need to know the price of the products, buyers' information and auctioning information. Livestock researchers and extension officers work hard for the farmers' benefit, but the current system shows that, to a great extent, farmers do not enjoy the expected services offered by extension officers as well as do not benefit from utilize results of research works.

3.3: Current channel and challenges of existing information flow methodology

From the study, 67.6% of livestock keepers depend on extension officers as their main source of information. The adopted model is, extension officers are supposed to digest the new knowledge and information acquired from research findings and other sources and to transmit it to livestock keepers, in a simple language that can be understood easily. Unfortunately, however, the results from the survey shows that, as many as 63.6% of extension officers face the same challenges of travelling long distances without reliable transport and sometimes possessing poor working tools or none at all. Another challenge is that of being overloaded with livestock keepers to serve. Researchers as well as extension officers have a key role to play in the development of the livestock industry. It is important therefore for all research findings to be communicated to all livestock stakeholders who will utilize them for their benefit.

The situation as it is now is that, livestock keepers find it difficult to get hold of the research findings because most of the researchers post their results in scientific journals and electronic media where livestock keepers cannot access. It is a problem that must be resolved. It can somehow be resolved if the research findings are posted in a central database, in a much simpler form so as to enable livestock keepers to easily access and understand the research findings by using their mobile phones. The results of the study show that 97.6% of farmers possess mobile phones. And 61.1% support the idea of using mobile phones as a means of accessing and sharing information. 94.8% believed that it will benefit them in using their mobile phones to obtain useful information from researchers, extension officers as well as from the markets.

It is also necessary to make improvements to extension services by providing extension officers with better transport facilities and to let them extensively use mobile phones to send consultative information to livestock keepers. Likewise, under this arrangement, livestock keepers will be able to easily communicate with extension officers to report any problems they face. 63.6% of extension officers said that they can use their mobile phones to offer such services.

3.4 Current Recording System

The results from the study indicate that 87.1% of the respondents maintain records for their operations, among them 51% use either exercise books or loose paper to record their data, while the visited dairy farms use specially designed forms to record their

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

farm's dairy cattle information. 92% of those who keep record of their operations, positively indicated that the data which they kept helped them very much in decision making.

Though much of the data kept was diseases issues. And for those who did not keep data (about 13%) said that they did not keep any data due to lack of resources, while some said they did not maintain any records because they did not have time and they thought it was unnecessary to keep records. But from the study done by [15] one of the recommendations was that the agriculture sector and other non-governmental agencies need to help farmers to increase their interest in keeping data by giving them tools that will facilitate recordkeeping.

From the survey it was found that the commonly ways of keeping data were, recording in exercise books, loose paper, and specially designed forms. And even by memorizing the data. Just over half (51%) of those who responded to the survey questions reported that they use exercise books and a few (13%) memorized the data while 36% used both ways. It was reported that some of the pocket record books, calendars are designed for initial record-keeping which need to be transferred and stored in the system that will allow easy analysis of data [5]. Thus in any system, this first level of data collection is very important as it is the key for having good information for decision making in the future. But, very unfortunately, most of the livestock keepers regard this crude way of record keeping as the end tool of data storage.

3.5 Challenges associated with current mechanism of storing Information

In response to the question raised to those livestock keepers who kept data, most of them indicated that they experienced some challenges/problems in their manner of storing data. Some of the challenges were, loss of record books, damage to the books and memory loss through forgetfulness. As for the dairy farms one of the challenges which was noticed was the excessive quantity of paper which was being used to make the recordings. They normally use specially designed forms. A separate form is used to record data for each livestock, on weekly basis.

Figure 4 shows the format of the recording form. Bearing in mind the big number of livestock kept in the dairy farms it is obvious that the number of record forms will be enormous and this might create problems during the process of analysing the data, especially if the data is analysed long periods after it was recorded. Under such circumstances the likelihood of getting an inaccurate analysis is real. [16] noted that manual record-keeping is usually time consuming, tedious to find important data and make decisions. Despite of the situation as it is now, it was observed that at least dairy farms have a better way of keeping records (Use of forms) compared to the way individual livestock keepers keep their records.

Figure 4: Forms for input records

Darrh [5] also identified that one of the challenges, even if farmers will be able to keep these books, remained to be in data analysis. Thus how were these records useful to farmers? He identified that the initial data, which is kept in exercise/note books is not very useful because it gives no means of comparing performance from year to year and in the face of varying challenges. When Analysis of the long-term records takes place it can help to pinpoint the weaknesses and strengths in the management program and help in identifying individual animals which fail to perform profitably. [17] reported that a Computerised record-keeping system improves the timeliness and accuracy of decision-making

3.6 Initiatives which have been taken.

From the interview with the Arumeru District Council it was learnt that the Government through the Ministry of Livestock and Fisheries has started to take some action towards improving the record keeping process. It has instructed extension officers to start using paper forms in which they are to record, in summary, all relevant data from the villagers whom they serve and forward the completed forms to the District Council where the data is recorded to a computer system. But even for the data so collected and stored, does not appear to be of help to the individual livestock keeper because he cannot easily access it and use it for the betterment of his livestock farm and to make it worse the data stored is in summary form as such it does not contain many details of useful need to the livestock keeper.

According to [5] stated that analysis of the long-term individual records can help to pinpoint weak areas in the management program and aid in identifying individual animals that fail to perform at profitable levels. But the only step that is missing is the way of taking those initial records and store them into a system for self-evaluation. And Good records are only beneficial if they are incorporated into management-making decisions. As a result records must be recorded accurately, analysed, and interpreted. But ICT gives solution to this where a

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Computerised record-keeping system improves the timeliness and accuracy of decision-making [17] . Therefore for easy storage, retrieval and analysis a computerized system is needed. Computers are powerful to the extent that they can analyse a huge amount of data within a short period of time.
Interviews, further, revealed that during extension service session and some of researches that are conducted depend on data collected by livestock keepers on a daily basis, from their operations. Researchers recommended that a central repository of data should be established so that livestock keepers can post day to day information about their animals, e.g. (Quantity of milk produced per day, livestock life log book, and check-up routines etc.) And this data would come in handy during research work and even in delivering online extension service. Taking advantage of the Government initiative of collecting summarized data from villages, the government can expand a bit by allowing farmers now to collect their data to the database for their individual assessment as shown in Figure 5.

Figure 5: Farmers is posting their daily farm records for easily analysis, and self-evaluation

4. HOW ICT HAS BEEN USED AS A TOOL FOR ACCESS AND BRIDGING INFORMATION FLOW GAP.

In simple terms ICT, refer to a growing assembly of technologies that are used to handle information and aid communication. It refers to various hardware and software for data collection, storage, processing, transmission and presentation of information in any format (i.e., voice, data, text and image). Some of the mostly used ICT hardware and software includes: the computers, the Internet, email, telephone, radio, television, video, digital cameras, etc. [1]. ICT can play a critical role in facilitating rapid, efficient, and cost effective knowledge management [18] [19] [20].
Notably the sources of agricultural knowledge include scientific research and indigenous knowledge. But after the creation, sourcing or accumulation of knowledge, the knowledge has to be disseminated to users to support the innovation process[21]. But it has been identified that there are inadequate number of extension officers to convey this knowledge to farmers as a result farmers remain uninformed.

One promising area to make extension service reach a large number of farmers is through technologies of mass communication (ICTs): cell phones, innovative community radio and television. ICT-based extension of agriculture brings wonderful opportunities and has the potential to facilitate the empowerment of the agriculture community. With the availability of ICTs proposal, to increase the number of extension workers may no longer be entirely valid. Because one extension officer can deliver information to a thousand farmers as ICTs is the faster media that can broadcast information to a large number of people regardless of distance provided that there is good communication infrastructure[22] [23].

Currently in Tanzania there are a number of ICT projects that have been implemented purposely for disseminating information to farmers, including: First Mile Project implemented between 2005 and 2009 by Agricultural Marketing Systems Development Program (AMSDP) in the northern and the southern highlands of Tanzania. The project promotes SMS, voice calls, and internet services to improve the availability. This project helps farmers to access timely and quality information on market prices, and it has improved communication amongst actors and local market [24].
Another project is the livestock Information Network and Knowledge System (LINKS) of the Global Livestock Collaborative Research Support Program (GL-CRSP). This project has strengthened the provision of regular livestock prices and volume information on most of the major livestock markets in Ethiopia, Kenya and Tanzania. Information is available on request via SMS text message system, email, World Space radio systems and on the internet. This provides a basis for livestock producers and traders to make informed marketing decisions[25].

Noted that 600 smallholder's farmers have benefited from a mobile agriculture solution program introduced by Vodacom, which is geared to improve agriculture. The project covers 30 markets in Tanzania, through mobile phones. Farmers located even in very remote areas are now accessing information [26].This service allows farmers to buy and sell crops and livestock.

Similarly TIGO has also invented a USSD system (TIGO kilimo) though it focused only in crop farmers. They gives a range of services including advice, marketing information and weather

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

information which currently operate in nine regions. Where a farmer have to choose the type of information he wants to access and the system reply by sending an SMS. It has also be a very easiest way that a farmers get services through his mobile phone.

FADECO (Family Alliance for Development and Cooperation), a local NGO reachable at (FRC 100.8 FM), operating in Karagwe District, Kagera region, uses radio for disseminating agricultural information and SMS to receive feedback and questions from farmers. In evaluating the impact of a distance education radio program in [27], it was also found that the farm school on radio with registered participants had a major impact on developing awareness, knowledge and changes in attitude.

Another radio launched June 28, 2013 known as Farm Radio International has strategies in helping women farmers in Tanzania to express their needs as women farmers as shown in Figure 6

Figure 6: Farm Radio International

Apart from that the Ministry of Livestock and Fisheries

Development and Ministry of Agriculture, Food Security and Cooperatives have websites (http://www.mifugouvuvi.go.tz/, and http://www.kilimo.go.tz/) as shown in Figure 7 and Figure 8 that provide various information related to livestock and agriculture crops to various stakeholders in the country and worldwide.

Figure 7: Website for Ministry of Agriculture, Food Security and Cooperatives

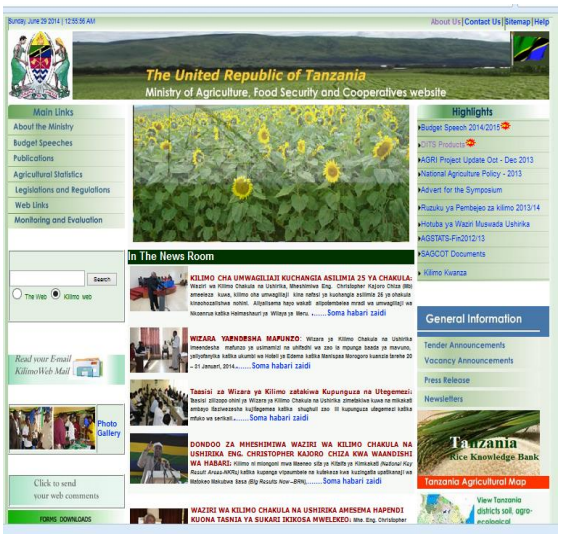


Figure 8: Website for Ministry of Livestock and Fisheries

In Africa, providers of extension services are urgently seeking for the best ways to support farmers in terms of information, technology, advice, and empowerment. As a result, there has been an emergence of innovative extension approaches including:

ESOKO, this is the most popular mobile based technology operating in various countries, including Tanzania. It has a way to meet the need of farmers having opportunities of selling their crops, making a decision on what type of crops to grow based on market trends, and reducing a number of middlemen at the time they want to sell their products by providing access to market information and information on the best crops to grow at a particular time [28]. They have a range of applications that a user choose what he needs as shown in Figure 9 and Figure 10.

Figure 9: ESOKO services

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

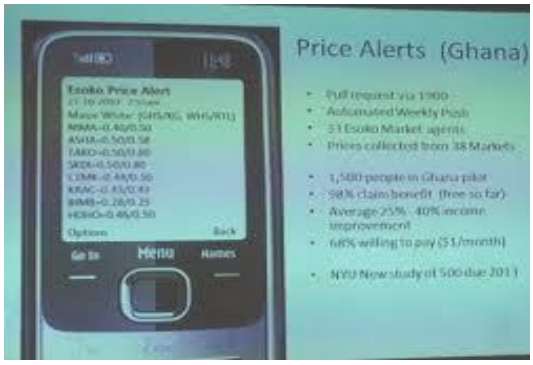


Figure 10: One of ESOKO services

M-Shamba
This is current being used by 4000 rice farmers in Kenya, is the applications accessible in both smart and low-end phones. Provides marketing, harvesting, production and weather information also it gives a room for farmers to share various topics with each other [29].

Mobile Agribiz: This is a web and SMS mobile application that assist farmers to decide when and how to plant the crops by providing weather information. This help a farmer to know the best type of the crops to grow per given weather conditions. Moreover it help to connect a farmer with a buyer by sending SMS with their phones numbers which are plotted into the map for buyers to see [30] .

KACE which operate in Kenya [31] [32] is the other rural-based market information system that allow farmers to link with buyers in different urban centres' [33]. It provides marketing and other agriculture related information[34].

Effects of ICT in providing extension and advisory services
There is few studies which have been conducted to analyse the impact of ICT in agriculture where a study conducted in Nigeria [35] shows that ICT have helped farmers to reduce searching cost as they use mobile phones which search a greater number of markets and sell in more. This was also confirmed by who conducted a study in Uganda where he found the expansion of the mobile phone coverage, results to increase in the proportion of the farmers who sold their crops in communities which are 20 miles away from district centres
Furthermore ICT have resulted to the increase in price transparency, thus if farmers will be provided with accurate and real-time price information will help them to be informed and sales based on what it is in the market [36] [37]. Also in providing extension services another study conducted by [36] in analysing a mobile based advisory service identify that there was an increase in the yearly income of the farmers for 37% after they received information on best way of faming.

General the use of ICT in dissemination of agricultural information has become very useful to farmers as:

a) It allows farmers to access timely information such as marketing, price etc., by helping to search for markets more efficiently and transparent that can reduce waste and empower smallholders in negotiation with wholesalers, traders and transport providers and link smallholders to distant markets and higher-end agricultural value chains.
b) ICT applications can improve advance warning of weather risks, pests and other environmental risks and provide timely, locally relevant information on how to respond to these.
c) Also, they can facilitate access to vital complementary services, particularly financial services

It have been established that farmers benefit by increasing production when they obtain knowledge on better practice of faming. Growing right crops at the right time that increase their income because they are being informed all the time on the current and future marketing trends. It is identified in [38] that the quality of information determines when information services are used and, further, what are the consequences of such usage. Therefore have the right information at the right time will improve their decision making.

5. CONCLUSION

From this study, it is obvious that the current status of information flow in the livestock sector is not satisfactory. It generally offers minimal contribution in completing the livestock chain/cycle and this leads to the deterioration of extension services to livestock keepers and to the overall performance of the sector. From what has been identified; ICT is considered to be a cutting edge solution in bridging this gap, where it gives an opportunity of conveying a message to millions users, provided there is good communication infrastructure. With the availability of ICTs the available number of extension workers may be fully utilized to reach many farmers.

It is noted that a number of farmers benefit from different studies that aim at minimizing the information gap. However, still there is no reliable direct link between a farmer and an extension officer most of these studies are mostly a one way communication (Giving out information without feedback), except to radio programs that can allow a farmer to call and ask questions. But a farmer has to be tuned on at that time and often are short time programs (30mins-1hr) where most of them cannot make it due to busy daily schedule.

Also giving advice on improved agricultural practices and farmer education, marketing information as separately services may not

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

necessarily lead to innovations and the desired increased productivity of smallholder agriculture [18] as farmers needs a complete extension services which can be achieved by integrating all services.

Also, it will be a good practice at a time a farmer is seeking for advice from expertise regarding his farm, for an expert to know exactly what is really going on at the farmer's farm (Informed advice). This can be achieved by referring to the data that has been collected from the farm. But using ICTs this can be achieved by having a system that allows a farmer to collect information regarding his farm and store it in a database. During advice session extension officer can browse through the database in order to formulate the advice that is relevant to what is happening in the farm (Give informed advice).
There are a several ICT tools that can be used to convey information to farmers, including mobile phones, computer, television, radios, to mention a few. The use of television and radios is somehow not reliable compared to the use of mobile phones and computer. Because the audience must know the exact time to tune in and listen given that most of these are short time programs, which is not enough for farmers to discuss all of their issues. But delivering of information through a computer and mobile phones can allow a farmer to get information at any time. Most of Tanzanian farmers have no computers. Thus the use of computer based communications is not the best media for conveying information. However, our study has found that a significant number of farmers possess mobile phones; which will be the best media for information delivery (Deliver information at the fingertips).

REFERENCE

1. Mekonnen, K.A.-O.a.D.A., *The Importance of ICTs in the Provision of Information for Improving Agricultural Productivity and Rural Incomes in Africa.* 2012(WP 2012-015:).
2. Hendriks, P., *Why share knowledge? The influence of ICT on the motivation for knowledge sharing.* Knowledge and process management, 1999. **6**(2): p. 91-100.
3. Blum, A., A. Lowengart-Aycicegi, and H. Magen, *Research Findings.*
4. DAVIS, K.E., *Agriculture and Climate Change: An Agenda for Negotiation in Copenhagen.* The Importance Role of Extension Systems, Focus, 2009. **16**.
5. Darrh Bullock, L.v.R., Jim Akers, and Alison Smith, *Record Keeping*. 2009, Kentucky.
6. Solomon Abegaz, K.A., Alemu Yami, Girma Abebe, and S.Z.a.A. Hirpa, *Records and Record Keeping*. CHAPTER THIRTEEN.
7. Rabin, M.I., *Apparatus and method for record keeping and information distribution*. 2003, Google Patents.
8. Batte, M.T., E. Jones, and G.D. Schnitkey, *Computer use by Ohio commercial farmers.* American Journal of Agricultural Economics, 1990. **72**(4): p. 935-945.
9. Frank Hartwich, M.O., Jeremiah Temu, *Tanzania's Red Meat Value Chain*. 2012, UNIDO (2012).
10. Masinde, M., A. Bagula, and N.J. Muthama. *The role of ICTs in downscaling and up-scaling integrated weather forecasts for farmers in sub-Saharan Africa*. in *Proceedings of the Fifth International Conference on Information and Communication Technologies and Development*. 2012. ACM.
11. Statistics, N.B.o., *Population Distribution by Age and Sex*. 2013.
12. Demiryurek, K., *Analysis of information systems and communication networks for organic and conventional hazelnut producers in the Samsun province of Turkey*. Agricultural systems, 2010. **103**(7): p. 444-452.
13. Daniel, E., et al., *ASSESSMENT OF AGRICULTURAL EXTENSION SERVICES IN TANZANIA. A CASE STUDY OF KYELA, SONGEA RURAL AND MOROGORO RURAL DISTRICTS.* 2013.
14. Pica-Ciamarra, U., et al., *Linking smallholders to livestock markets: Combining market and household survey data in Tanzania.* 2011.
15. Tham-Agyekum, E.K., P. Appiah, and F. Nimoh, *Assessing Farm Record Keeping Behaviour among Small-Scale Poultry Farmers in the Ga East Municipality.* Journal of Agricultural Science (1916-9752), 2010. **2**(4).
16. Jeyabalan, V., *Individual cow recording and analysis system for small scale dairy farmers in Malaysia.* International Journal of Computer Applications, 2010. **8**(11).
17. Jofre-Giraudo, E., D.H. Streeter, and W. Lazarus, *The impact of computer information systems on dairy farm management decisions.* Agribusiness, 1990. **6**(5): p. 463-474.
18. Asenso-Okyere, K. and D.A. Mekonnen, *The importance of ICTs in the provision of information for improving agricultural productivity and rural incomes in Africa.* African Human Development Report. UNDP Sponsored research Series, 2012.
19. Chapman, R. and T. Slaymaker, *ICTs and Rural Development: Review of the Literature, Current.* 2002.
20. Huyer, S. and S. Mitter, *ICTs, globalisation and poverty reduction: Gender dimensions of the knowledge society.* Kampala (Uganda): Comisión de Ciencia y Tecnología para el Desarrollo (Naciones Unidas), Junta Consultiva sobre Cuestiones de Género. Puede consultarse en http://gab. wigsat. org/policy. htm, 2003.
21. UNDP, *Promoting ICT based agricultural knowledge management to increase production and productivity of smallholder farmers in Ethiopia.* 2013.
22. Megwa, E.R., *Bridging the digital divide: Community radio's potential for extending information and communication technology benefits to poor rural communities in South Africa*. The Howard Journal of Communications, 2007. **18**(4): p. 335-352.
23. Bobbili, R., et al., *Radio farm forum and Afronet: learning from successful ICT projects in Zambia*. Zambia 21F, 2006.
24. Lightfoot, C., et al., *The first mile project in Tanzania: linking smallholder farmers to markets using modern communication technology.* Mountain Research and Development, 2008. **28**(1): p. 13-17.
25. LINKS. *Livestock Information Network Knowledge System*. 2014; Available from: http://www.lmistz.net/Pages/Public/Home.aspx.
26. AMSDP, *Agricultural Marketing Systems Development Programme (AMSDP)*. 2009.
27. Sasidhar, P., et al., *Evaluation of a distance education radio farm school programme in India: implications for scaling up.* Outlook on AGRICULTURE, 2011. **40**(1): p. 89-96.
28. David-West, O., *Esoko Networks: facilitating agriculture through technology*. 2011, New York: United Nations Development Project.
29. mshamba. *Mshamba services.* 2014; Available from: http://www.mshamba.net/.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

30. Agribiz. *Web and Mobile app that connects farmers to the markets* 2014; Available from: http://www.synapsecenter.org/giresse-mila-lokwa.
31. Mayaka, P.K., *A Knowledge Management System for Horticulture Farming in Kenya*. 2013, University of Nairobi.
32. Islam, M.S., *Creating opportunity by connecting the unconnected: mobile phone based agriculture market information service for farmers in Bangladesh.* 2011.
33. Muriithi, A.G., E. Bett, and S.A. Ogaleh. *Information Technology for Agriculture and Rural Development in Africa: Experiences from Kenya*. in *Conference on International Research on Food Security, Natural Resource Management and Rural Development*. 2009.
34. Saravanan, R., *ICTs for Agricultural Extension: Global Experiments, Innovations and Experiences*. 2010: New India Publishing.
35. Aker, J.C., *Does digital divide or provide? The impact of cell phones on grain markets in Niger.* Center for Global Development Working Paper, 2008. **154**.
36. Grimshaw, D.J. and S. Kala, *Strengthening Rural Livelihoods: The impact of information and communication technologies in Asia*. 2011: IDRC.
37. Baumüller, H., *Facilitating agricultural technology adoption among the poor: The role of service delivery through mobile phones.* 2012, ZEF Working Paper Series.
38. Srinivasan, J. *The role of trustworthiness in information service usage: The case of Parry information kiosks, Tamil Nadu, India*. in *Information and Communication Technologies and Development, 2007. ICTD 2007. International Conference on*. 2007. IEEE.

IJCSIS AUTHORS' & REVIEWERS' LIST

Assist Prof (Dr.) M. Emre Celebi, Louisiana State University in Shreveport, USA
Dr. Lam Hong Lee, Universiti Tunku Abdul Rahman, Malaysia
Dr. Shimon K. Modi, Director of Research BSPA Labs, Purdue University, USA
Dr. Jianguo Ding, Norwegian University of Science and Technology (NTNU), Norway
Assoc. Prof. N. Jaisankar, VIT University, Vellore,Tamilnadu, India
Dr. Amogh Kavimandan, The Mathworks Inc., USA
Dr. Ramasamy Mariappan, Vinayaka Missions University, India
Dr. Yong Li, School of Electronic and Information Engineering, Beijing Jiaotong University, P.R. China
Assist. Prof. Sugam Sharma, NIET, India / Iowa State University, USA
Dr. Jorge A. Ruiz-Vanoye, Universidad Autónoma del Estado de Morelos, Mexico
Dr. Neeraj Kumar, SMVD University, Katra (J&K), India
Dr Genge Bela, "Petru Maior" University of Targu Mures, Romania
Dr. Junjie Peng, Shanghai University, P. R. China
Dr. Ilhem LENGLIZ, HANA Group - CRISTAL Laboratory, Tunisia
Prof. Dr. Durgesh Kumar Mishra, Acropolis Institute of Technology and Research, Indore, MP, India
Dr. Jorge L. Hernández-Ardieta, University Carlos III of Madrid, Spain
Prof. Dr.C.Suresh Gnana Dhas, Anna University, India
Dr Li Fang, Nanyang Technological University, Singapore
Prof. Pijush Biswas, RCC Institute of Information Technology, India
Dr. Siddhivinayak Kulkarni, University of Ballarat, Ballarat, Victoria, Australia
Dr. A. Arul Lawrence, Royal College of Engineering & Technology, India
Dr. Wongyos Keardsri, Chulalongkorn University, Bangkok, Thailand
Dr. Somesh Kumar Dewangan, CSVTU Bhilai (C.G.)/ Dimat Raipur, India
Dr. Hayder N. Jasem, University Putra Malaysia, Malaysia
Dr. A.V.Senthil Kumar, C. M. S. College of Science and Commerce, India
Dr. R. S. Karthik, C. M. S. College of Science and Commerce, India
Dr. P. Vasant, University Technology Petronas, Malaysia
Dr. Wong Kok Seng, Soongsil University, Seoul, South Korea
Dr. Praveen Ranjan Srivastava, BITS PILANI, India
Dr. Kong Sang Kelvin, Leong, The Hong Kong Polytechnic University, Hong Kong
Dr. Mohd Nazri Ismail, Universiti Kuala Lumpur, Malaysia
Dr. Rami J. Matarneh, Al-isra Private University, Amman, Jordan
Dr Ojesanmi Olusegun Ayodeji, Ajayi Crowther University, Oyo, Nigeria
Dr. Riktesh Srivastava, Skyline University, UAE
Dr. Oras F. Baker, UCSI University - Kuala Lumpur, Malaysia
Dr. Ahmed S. Ghiduk, Faculty of Science, Beni-Suef University, Egypt
and Department of Computer science, Taif University, Saudi Arabia
Dr. Tirthankar Gayen, IIT Kharagpur, India
Dr. Huei-Ru Tseng, National Chiao Tung University, Taiwan

CALL FOR PAPERS
International Journal of Computer Science and Information Security

IJCSIS 2014
ISSN: 1947-5500
http://sites.google.com/site/ijcsis/

International Journal Computer Science and Information Security, IJCSIS, is the premier scholarly venue in the areas of computer science and security issues. IJCSIS 2011 will provide a high profile, leading edge platform for researchers and engineers alike to publish state-of-the-art research in the respective fields of information technology and communication security. The journal will feature a diverse mixture of publication articles including core and applied computer science related topics.

Authors are solicited to contribute to the special issue by submitting articles that illustrate research results, projects, surveying works and industrial experiences that describe significant advances in the following areas, but are not limited to. Submissions may span a broad range of topics, e.g.:

Track A: Security

Access control, Anonymity, Audit and audit reduction & Authentication and authorization, Applied cryptography, Cryptanalysis, Digital Signatures, Biometric security, Boundary control devices, Certification and accreditation, Cross-layer design for security, Security & Network Management, Data and system integrity, Database security, Defensive information warfare, Denial of service protection, Intrusion Detection, Anti-malware, Distributed systems security, Electronic commerce, E-mail security, Spam, Phishing, E-mail fraud, Virus, worms, Trojan Protection, Grid security, Information hiding and watermarking & Information survivability, Insider threat protection, Integrity
Intellectual property protection, Internet/Intranet Security, Key management and key recovery, Language-based security, Mobile and wireless security, Mobile, Ad Hoc and Sensor Network Security, Monitoring and surveillance, Multimedia security ,Operating system security, Peer-to-peer security, Performance Evaluations of Protocols & Security Application, Privacy and data protection, Product evaluation criteria and compliance, Risk evaluation and security certification, Risk/vulnerability assessment, Security & Network Management, Security Models & protocols, Security threats & countermeasures (DDoS, MiM, Session Hijacking, Replay attack etc,), Trusted computing, Ubiquitous Computing Security, Virtualization security, VoIP security, Web 2.0 security, Submission Procedures, Active Defense Systems, Adaptive Defense Systems, Benchmark, Analysis and Evaluation of Security Systems, Distributed Access Control and Trust Management, Distributed Attack Systems and Mechanisms, Distributed Intrusion Detection/Prevention Systems, Denial-of-Service Attacks and Countermeasures, High Performance Security Systems, Identity Management and Authentication, Implementation, Deployment and Management of Security Systems, Intelligent Defense Systems, Internet and Network Forensics, Large-scale Attacks and Defense, RFID Security and Privacy, Security Architectures in Distributed Network Systems, Security for Critical Infrastructures, Security for P2P systems and Grid Systems, Security in E-Commerce, Security and Privacy in Wireless Networks, Secure Mobile Agents and Mobile Code, Security Protocols, Security Simulation and Tools, Security Theory and Tools, Standards and Assurance Methods, Trusted Computing, Viruses, Worms, and Other Malicious Code, World Wide Web Security, Novel and emerging secure architecture, Study of attack strategies, attack modeling, Case studies and analysis of actual attacks, Continuity of Operations during an attack, Key management, Trust management, Intrusion detection techniques, Intrusion response, alarm management, and correlation analysis, Study of tradeoffs between security and system performance, Intrusion tolerance systems, Secure protocols, Security in wireless networks (e.g. mesh networks, sensor networks, etc.), Cryptography and Secure Communications, Computer Forensics, Recovery and Healing, Security Visualization, Formal Methods in Security, Principles for Designing a Secure Computing System, Autonomic Security, Internet Security, Security in Health Care Systems, Security Solutions Using Reconfigurable Computing, Adaptive and Intelligent Defense Systems, Authentication and Access control, Denial of service attacks and countermeasures, Identity, Route and

Location Anonymity schemes, Intrusion detection and prevention techniques, Cryptography, encryption algorithms and Key management schemes, Secure routing schemes, Secure neighbor discovery and localization, Trust establishment and maintenance, Confidentiality and data integrity, Security architectures, deployments and solutions, Emerging threats to cloud-based services, Security model for new services, Cloud-aware web service security, Information hiding in Cloud Computing, Securing distributed data storage in cloud, Security, privacy and trust in mobile computing systems and applications, **Middleware security & Security features:** middleware software is an asset on
its own and has to be protected, interaction between security-specific and other middleware features, e.g., context-awareness, **Middleware-level security monitoring and measurement:** metrics and mechanisms for quantification and evaluation of security enforced by the middleware, **Security co-design:** trade-off and co-design between application-based and middleware-based security, **Policy-based management:** innovative support for policy-based definition and enforcement of security concerns, **Identification and authentication mechanisms:** Means to capture application specific constraints in defining and enforcing access control rules, **Middleware-oriented security patterns:** identification of patterns for sound, reusable security, **Security in aspect-based middleware:** mechanisms for isolating and enforcing security aspects, **Security in agent-based platforms:** protection for mobile code and platforms, Smart Devices: Biometrics, National ID cards, Embedded Systems Security and TPMs, RFID Systems Security, Smart Card Security, Pervasive Systems: Digital Rights Management (DRM) in pervasive environments, Intrusion Detection and Information Filtering, Localization Systems Security (Tracking of People and Goods), Mobile Commerce Security, Privacy Enhancing Technologies, Security Protocols (for Identification and Authentication, Confidentiality and Privacy, and Integrity), Ubiquitous Networks: Ad Hoc Networks Security, Delay-Tolerant Network Security, Domestic Network Security, Peer-to-Peer Networks Security, Security Issues in Mobile and Ubiquitous Networks, Security of GSM/GPRS/UMTS Systems, Sensor Networks Security, Vehicular Network Security, Wireless Communication Security: Bluetooth, NFC, WiFi, WiMAX, WiMedia, others

This Track will emphasize the design, implementation, management and applications of computer communications, networks and services. Topics of mostly theoretical nature are also welcome, provided there is clear practical potential in applying the results of such work.

Track B: Computer Science

Broadband wireless technologies: LTE, WiMAX, WiRAN, HSDPA, HSUPA, Resource allocation and interference management, Quality of service and scheduling methods, Capacity planning and dimensioning, Cross-layer design and Physical layer based issue, Interworking architecture and interoperability, Relay assisted and cooperative communications, Location and provisioning and mobility management, Call admission and flow/congestion control, Performance optimization, Channel capacity modeling and analysis, Middleware Issues: Event-based, publish/subscribe, and message-oriented middleware, Reconfigurable, adaptable, and reflective middleware approaches, Middleware solutions for reliability, fault tolerance, and quality-of-service, Scalability of middleware, Context-aware middleware, Autonomic and self-managing middleware, Evaluation techniques for middleware solutions, Formal methods and tools for designing, verifying, and evaluating, middleware, Software engineering techniques for middleware, Service oriented middleware, Agent-based middleware, Security middleware, Network Applications: Network-based automation, Cloud applications, Ubiquitous and pervasive applications, Collaborative applications, RFID and sensor network applications, Mobile applications, Smart home applications, Infrastructure monitoring and control applications, Remote health monitoring, GPS and location-based applications, Networked vehicles applications, Alert applications, Embeded Computer System, Advanced Control Systems, and Intelligent Control : Advanced control and measurement, computer and microprocessor-based control, signal processing, estimation and identification techniques, application specific IC's, nonlinear and adaptive control, optimal and robot control, intelligent control, evolutionary computing, and intelligent systems, instrumentation subject to critical conditions, automotive, marine and aero-space control and all other control applications, Intelligent Control System, Wiring/Wireless Sensor, Signal Control System. Sensors, Actuators and Systems Integration : Intelligent sensors and actuators, multisensor fusion, sensor array and multi-channel processing, micro/nano technology, microsensors and microactuators, instrumentation electronics, MEMS and system integration, wireless sensor, Network Sensor, Hybrid

Sensor, Distributed Sensor Networks. Signal and Image Processing : Digital signal processing theory, methods, DSP implementation, speech processing, image and multidimensional signal processing, Image analysis and processing, Image and Multimedia applications, Real-time multimedia signal processing, Computer vision, Emerging signal processing areas, Remote Sensing, Signal processing in education. Industrial Informatics: Industrial applications of neural networks, fuzzy algorithms, Neuro-Fuzzy application, bioInformatics, real-time computer control, real-time information systems, human-machine interfaces, CAD/CAM/CAT/CIM, virtual reality, industrial communications, flexible manufacturing systems, industrial automated process, Data Storage Management, Harddisk control, Supply Chain Management, Logistics applications, Power plant automation, Drives automation. Information Technology, Management of Information System : Management information systems, Information Management, Nursing information management, Information System, Information Technology and their application, Data retrieval, Data Base Management, Decision analysis methods, Information processing, Operations research, E-Business, E-Commerce, E-Government, Computer Business, Security and risk management, Medical imaging, Biotechnology, Bio-Medicine, Computer-based information systems in health care, Changing Access to Patient Information, Healthcare Management Information Technology. Communication/Computer Network, Transportation Application : On-board diagnostics, Active safety systems, Communication systems, Wireless technology, Communication application, Navigation and Guidance, Vision-based applications, Speech interface, Sensor fusion, Networking theory and technologies, Transportation information, Autonomous vehicle, Vehicle application of affective computing, Advance Computing technology and their application : Broadband and intelligent networks, Data Mining, Data fusion, Computational intelligence, Information and data security, Information indexing and retrieval, Information processing, Information systems and applications, Internet applications and performances, Knowledge based systems, Knowledge management, Software Engineering, Decision making, Mobile networks and services, Network management and services, Neural Network, Fuzzy logics, Neuro-Fuzzy, Expert approaches, Innovation Technology and Management : Innovation and product development, Emerging advances in business and its applications, Creativity in Internet management and retailing, B2B and B2C management, Electronic transceiver device for Retail Marketing Industries, Facilities planning and management, Innovative pervasive computing applications, Programming paradigms for pervasive systems, Software evolution and maintenance in pervasive systems, Middleware services and agent technologies, Adaptive, autonomic and context-aware computing, Mobile/Wireless computing systems and services in pervasive computing, Energy-efficient and green pervasive computing, Communication architectures for pervasive computing, Ad hoc networks for pervasive communications, Pervasive opportunistic communications and applications, Enabling technologies for pervasive systems (e.g., wireless BAN, PAN), Positioning and tracking technologies, Sensors and RFID in pervasive systems, Multimodal sensing and context for pervasive applications, Pervasive sensing, perception and semantic interpretation, Smart devices and intelligent environments, Trust, security and privacy issues in pervasive systems, User interfaces and interaction models, Virtual immersive communications, Wearable computers, Standards and interfaces for pervasive computing environments, Social and economic models for pervasive systems, Active and Programmable Networks, Ad Hoc & Sensor Network, Congestion and/or Flow Control, Content Distribution, Grid Networking, High-speed Network Architectures, Internet Services and Applications, Optical Networks, Mobile and Wireless Networks, Network Modeling and Simulation, Multicast, Multimedia Communications, Network Control and Management, Network Protocols, Network Performance, Network Measurement, Peer to Peer and Overlay Networks, Quality of Service and Quality of Experience, Ubiquitous Networks, Crosscutting Themes – Internet Technologies, Infrastructure, Services and Applications; Open Source Tools, Open Models and Architectures; Security, Privacy and Trust; Navigation Systems, Location Based Services; Social Networks and Online Communities; ICT Convergence, Digital Economy and Digital Divide, Neural Networks, Pattern Recognition, Computer Vision, Advanced Computing Architectures and New Programming Models, Visualization and Virtual Reality as Applied to Computational Science, Computer Architecture and Embedded Systems, Technology in Education, Theoretical Computer Science, Computing Ethics, Computing Practices & Applications

Authors are invited to submit papers through e-mail ijcsiseditor@gmail.com. Submissions must be original and should not have been published previously or be under consideration for publication while being evaluated by IJCSIS. Before submission authors should carefully read over the journal's Author Guidelines, which are located at http://sites.google.com/site/ijcsis/authors-notes .

ISSN 1947 5500
http://sites.google.com/site/ijcsis/