International Journal of Computer Applications Technology and Research
Volume 5– Issue 5, 238 - 240, 2016, ISSN:- 2319–8656

Cryptographic Hash Function using Cellular Automata

Mohamed Mahmoud ElRakaiby

Communications and Electronics Department
Alexandria University
Alexandria, Egypt

Abstract: In this paper we make use of statistical properties of applying elementary cellular automata on a block of bits to generate a fixed size digest of that block to use it as hash function which can be use in different cryptographic applications.

Keywords: file digest; hash; one way function; cellular automata;

1. INTRODUCTION

Hash functions **H** plays an important role in different cryptographic applications which rely on hash function only or combine it with other cryptographic standards to produce a certain protocol suite a desired application

Data integrity is the most important application of hash functions. By using the message digests generated by a cryptographic hash function a system administrator can detect unauthorized changes in files.

In this paper we will introduce a new type of Hash functions using Cellular Automata which will be explained briefly as well

2. Hash Functions

Hash functions are mathematical computations that take in a relatively arbitrary amount of data as input and produce an output of fixed size. The output is always the same when given the same input (Figure 1). The inputs to a hash function are typically called messages, and the outputs are often referred to as message digests

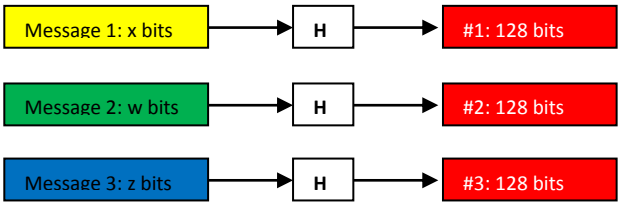


Figure 1. Hash Function Basic Description

All hash functions have the property that it is impossible to determine the input knowing only the output

Currently we have two main popular hash functions Message Digest 5 MD5 [1] which produce 128 bit digest for any input length and Secure Hash Algorithms SHA [2] which produces variety of digest sizes according to used standard (160 – 512) bits

Since two distinct messages are extremely unlikely to generate identical message digests, one can use this property of cryptographic hash functions to detect when a message has been altered. If one takes a binary file and computes a digest of the file, one can record this baseline digest. In the future, the digest can be recomputed on the file. If the new digest differs from the original baseline digest, then one can be assured that the file has been altered in some way

3. Cellular Automata

3.1 Elementary Cellular Automata

A cellular automata (CA) consists of a regular grid of cells, each in one of a finite number of states [3], such as on and off. The grid can be in any finite number of dimensions. For each cell, a set of cells called its neighborhood is defined relative to the specified cell. An initial state (time t = 0) is selected by assigning a state for each cell. A new generation is created (at t=t+1), according to some fixed rule that determines the new state of each cell in terms of the current state of the cell and the states of the cells in its neighborhood. Typically, the rule for updating the state of cells is the same for each cell and does not change over time, and is applied to the whole grid simultaneously

The simplest cellular automata system is one dimensional with two possible states per cell which we call it Elementary Cellular Automata with two adjacent neighbors per cell and according to cell status (Figure 2), neighbors status at time t=T , we get the cell status at time t=T+1 controlled by transition rule

Figure 2. Cell and two neighbors

Cellular Automata also can be two dimensional [4] or three dimensional [5] which can be used in different applications but in this paper we use the simplest model of cellular automata

Figure 4 shows example of Rule 90 (Figure 3) in elementary cellular automata with 2 neighbors per cell (90=01011010), also Figure shows graphical representation of applying Rule 90 on 1001011000000100001001000000011 for 16 iteration

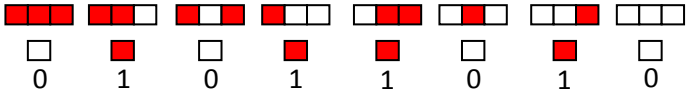


Figure 3. Rule 90

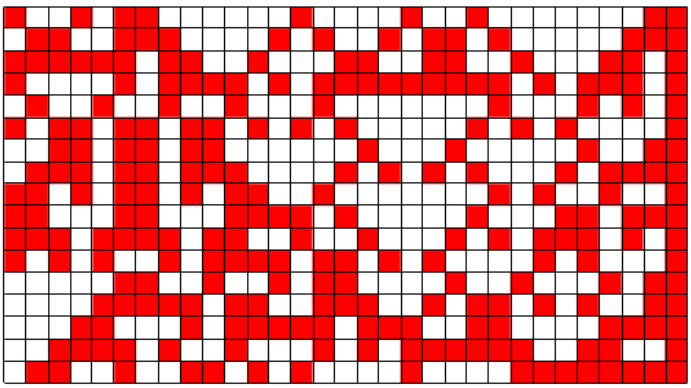


Figure 4. 16 Iterations using Rule 90

3.2 Cell Neighbors and Rule expansion

As we learnt so far that rule can be represented as the decimal equivalent of truth table output whose inputs are the all combinations of cell-neighbors states , it is easy in case of three neighbors rules

In our proposal in paper we will expand the neighbors to be six neighbors three on left of cell and three on right of cell (Figure 5) which will produce 128 combinations

Figure 5. Cell and 6 neighbors

Rule is composed from bits of the right column of table 1 which represent the result of truth table of all combinations of cell and its neighbors

Table 1. Truth Table of Cell and 6 neighbors

N^{-3}	N^{-2}	N^{-1}	C^{t}	N^{1}	N^{2}	N^{3}	C^{t+1}
0	0	0	0	0	0	0	C^{127}
0	0	0	0	0	0	1	C^{126}
0	0	0	0	0	1	0	C^{125}
0	0	0	0	0	1	1	C^{124}
0	0	0	0	1	0	0	C^{123}
0	0	0	0	1	0	1	C^{122}
			.				
			.				
1	1	1	1	0	1	1	C^{4}
1	1	1	1	1	0	0	C^{3}
1	1	1	1	1	0	1	C^{2}
1	1	1	1	1	1	0	C^{1}
1	1	1	1	1	1	1	C^{0}

4. Proposed System

We propose a main building block which the file to be hashed should be multiples of its size which is 8 * 128 bit

In order to maximize the diffusion, we first take each 128 bit block to be processed as shown in below block using elementary Cellular Automata using r=3

We use each block as a rule for elementary cellular automata for another block of same size then, use the result as next rule for next block and so on till finishing the remaining blocks of bits (7 blocks) .This process result of one block of 128 bits for each input block as shown in Figure 6.

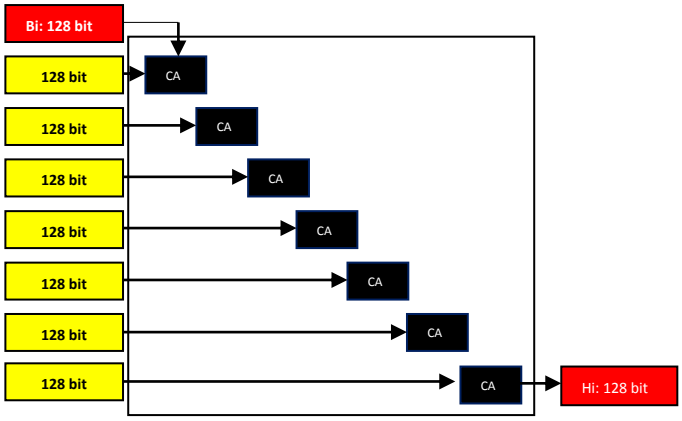


Figure 6. Initial Hashing of one block

Repeating this process for all 8 blocks can be presented as 8*8 initial hashing block with 8 inputs and 8 outputs (Figure 7).

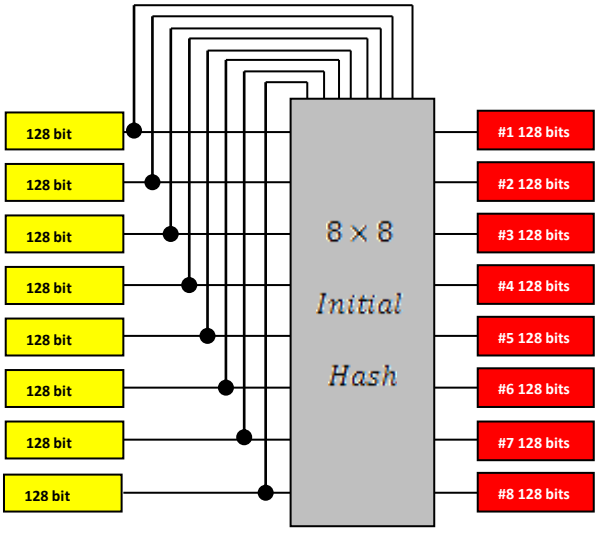


Figure 7. Initial Hashing 8 blocks

At that point, we produced 8*128 bit blocks as a result of the same size input so; we need to proceed in another step which is reduction step by processing each two successive blocks of 128 bits to produce only one block of 128 bit as shown in Figure 8.

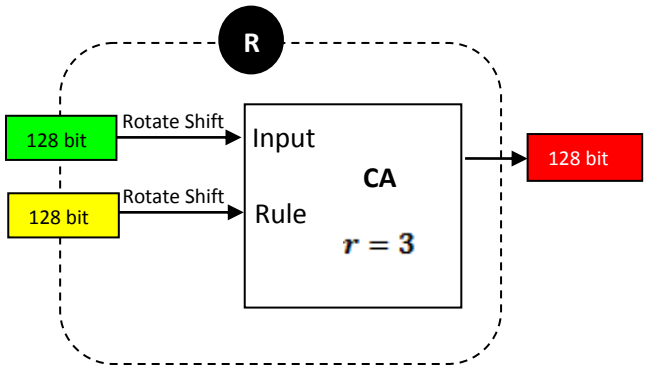


Figure 8. Single reduction round

Applying the reduction process on 8*128 bit blocks will result in 4*128 bit blocks and so on till we reach only one block of 128 (Figure 9) which will be considered as the digest of 8*128 bit blocks

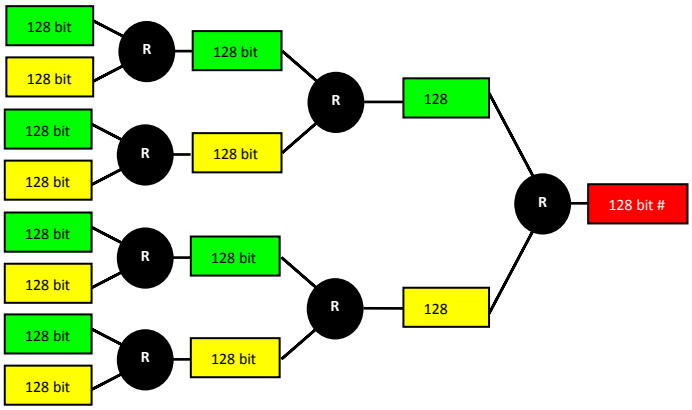


Figure 9. Reduction step of 8 x 128 bits to 128 bits

Last step is applying the same reduction process for resulting digests of building blocks for larger file sizes till we finally get fixed size hash for the whole file which is 128bit

Figure 10 is summary of hashing 4048 bits file

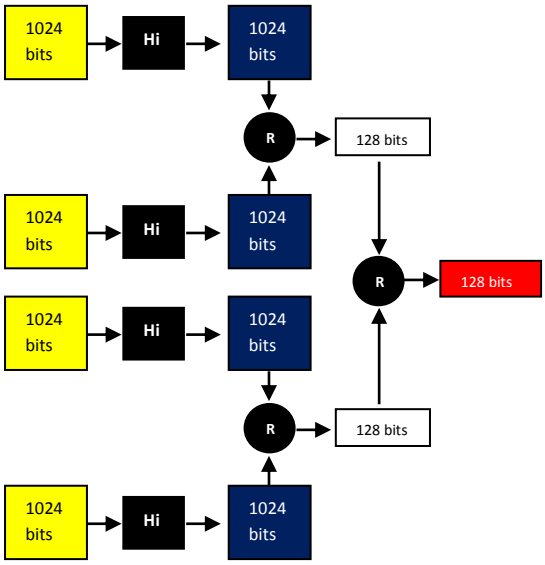


Figure 10. Summary of hashing 4048 bits file

5. Conclusion

We made use of cellular automata to prove that it can be used as a message digest function which produces a fixed size digest of a variable sizes files .It is sensitive to input file variation as shown in Figure 11

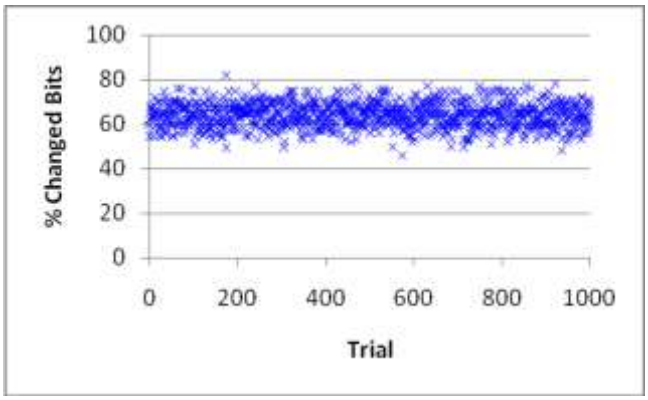


Figure 11. Changed bits after changing input by one bit

6. ACKNOWLEDGMENTS

Deep thanks to Prof. Hassan ElKamchouchi and Dr Fatma Ahmed for their continuous and endless support

7. REFERENCES

[1] Rivest RL (1991) The MD4 message-digest algorithm. Crypto, LNCS 537:303–311

[2] Secure Hash Standard (SHS), Federal Information Processing Standards Publication FIPS, PUB 180-4.

[3] Wolfram, Stephen (2002). A New Kind of Science. Wolfram Media.

[4] Norman H. Packard 1 and Stephen Wolfram, Two-Dimensional Cellular Automata , Journal of Statistical Physics, Vol. 38, Nos. 5/6, 1985

[5] R.W. Gerling , Classification of three-dimensional cellular automata , Physica A: Statistical Mechanics and its Applications , Volume 162, Issue 2, 1 January 1990, Pages 187-195

International Journal of Computer Applications Technology and Research
Volume 5– Issue 5, 241 - 244, 2016, ISSN:- 2319–8656

Energy Conservation in a Smart Home via an Embedded Platform

Rahul Sivaram
MIT College of Engineering
Ex-Serviceman Colony
Pune, India

Shashank Kumar
MIT College of Engineering
Ex-Serviceman Colony
Pune, India

Sahil Budhwani
MIT College of Engineering
Ex-Serviceman Colony
Pune, India

Vikash Kumar
MIT College of Engineering
Ex-Serviceman Colony
Pune, India

Abstract: Home automation is a popular field of application which combines the concepts of embedded systems and pervasive computing to control a living space for modern times. A smart home system aims to free up the user to carry out his or her daily tasks without worrying on the trivial aspects of personal home management, while at the same time providing the user with more control of the home than ever before. The implementation comprises of a large number of sensors which can be used to control or monitor objects distributed in three-dimensional space around the house. The sensors can be specialized in measuring temperature, humidity, pressure, light, noise, dust air, and upon intelligently computed triggers, compel the system to perform a specific task, or a set of tasks. In this project, a solution to transform a normal house into a smart, autonomous house in order to reduce the energy consumption of the household is proposed. This can be realized with the help of wired sensor networks where we control and interface with electronic appliances via a Raspberry Pi, an embedded development platform. The embedded platform can be controlled remotely to turn switches on and off, and eventually aims to become as autonomous as possible via the help of a Bayesian Network. The base workings of the project are aimed at being as scalable as possible, to be able to potentially fit into industrial as well as office environments in the future.

Keywords: Energy Conservation, Embedded Platform, Bayesian Belief Network, Pervasive Computing.

1. INTRODUCTION

In the 21st century, the advent of the internet and advances in the power and miniaturization of computing devices has made it possible to implement computers wherever we can imagine. The range of possibilities has given birth to the idea of 'Home Automation', in which computers and algorithms can be given the responsibilities of operating a household, with minimal Human interaction needed.

With the rise in popularity of cheap, embedded computers, such as the Raspberry Pi [8] and the Beaglebone Black computing has never been more accessible to the masses. It is proposed in this paper to take advantage of the small form factor, energy efficiency, computing power and GPIO capabilities, to interface the embedded platform with appliances and electrical sockets found in a household. These platforms can then be programmed in a way to tailor to the specific functions of the appliance it is connected to, essentially adding a level of intelligence to the otherwise "dumb" appliance. The addition of wireless modules to the embedded board adds the provision to wirelessly network the board, giving rise to the possibilities of an interconnected, intelligent home. The optimization of this wireless network could bring about a level of efficiency to a house/office space/industrial complex (depending on the appropriate scale factor) which could bring about tremendous gains in automation of menial tasks, interconnectivity of the application space, and energy usage savings.

Electricity is a precious resource which should be utilized carefully, especially in the modern era where energy requirements are rising at an alarming rate. It has become a common agony for citizens of urban spaces to face the ill effects of a poorly structured, operated and maintained energy grid. These include problems such as load shedding, and even complete blackouts. Some of the areas receive only an hour of electricity a day. As we are restricted on the front of producing electricity, this calls for a smart way to use electricity in our living spaces to create an energy efficient future. This is where our project comes in. We propose making our homes operationally seamless so that not only do we not have to worry about explicitly operating our home appliances, we can be assured the home space is functioning as efficiently and safely as possible [7]. A scenario in which every appliance turns on automatically whenever a person enters the space relevant to the context of the appliance, depending upon interlinked sensors and historical data which will act as the training set for our system. Since we will be working on limited resources we shall consider only a few appliances however the scope can be widened to accommodate various devices and so that we can obtain

optimized metrics, and perform critical analytics and improve the learning algorithm used in the project to ensure optimal functionality [5].

The paper proposes the use of a Bayesian Network as the principle machine learning algorithm to be used. The selection of this algorithm is due to the fact it is an extremely powerful tool to model uncertainty in a situation. The algorithm does this through the data present in a Joint Probability Distribution, in which the probabilities of events are co-related, and are used to teach a machine to come to a decision about an event when there are a set of actions the machine is allowed to take in the case of the event. This principle is used to model the uncertain way a home environment operates. even the average two bedroom home, in the perspective of the appliances the home may contain, the living and usage patterns of the occupants of the home, the considerations of users of the home space who may not be permanent residents, but rather frequent visitors such as family and friends, and non-frequent users such as delivery men, semi-pertinent users such as pets, makes modeling a deterministic "trigger" mechanism for the appliances of the home quite difficult, challenging the placement of this particular computational problem in the appropriate complexity class. Using the optimally designed Bayesian Network and historical data collected from different input points into the system, it is proposed to model this "trigger" mechanism [3].

2. RELATED WORK

In a paper described by the Silviu Folea Automation Department Technical University of Cluj-Napoca Cluj-Napoca, Romania[1], the fundamental concepts which describe a framework that can be applied to develop a smart home, and can be tweaked/expanded upon as necessary. It also provides an introduction to the software "LabVIEW", which can be used to capture and interpret data from various presence sensors.

Alsheikh et al[2] in their paper take into account the framework of Wireless Sensor Networks (WSNs) and introduces pertinent topics such as issues with wireless networks, organization of WSNs, and the specifications of WSNs. The paper also discusses various machine learning algorithms (Supervised and Unsupervised) which can be applied to WSNs to optimize their functioning as well as bring a degree of autonomy to them.

A paper authored by Nancy A. Roberts for the US Air force Research Laboratory[3], introduces the implementation of the concept of a Bayesian Network to model scenarios in a smart living space for security purposes, illustrated in the paper via a simulated "Door Break-in", in which complete knowledge of the domain is unknown. More specifically, it proposes how a Bayesian Network could use historical data to attempt to predict future instances of the "Door Break-in".

A paper by Dr. Leslie Haddon [5] deals with the less technical, but nevertheless important topics regarding the aspects of home automation such as the human attitudes towards the concept of home automation, as well as how the automation of another system such as electricity production and traffic management would be applicable

3. SYSTEM ARCHITECTURE

As can be seen in the FIG 3.1, we propose a twofold approach to interact with the embedded platform: one is via a smartphone application which can access a dynamic webpage encoded in a scripting language such as PHP, which is hosted by a server being run in the embedded platform, implemented using the appropriate framework such as Apache or Node.js.

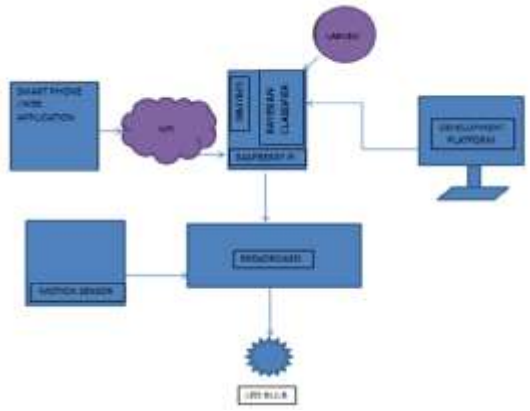


FIG 3.1: SYSTEM ARCHITECTURE

This approach demonstrates the act of explicit Human Computer Interaction (eHCI). The second approach is to have the system continuously monitor surrounding activity through a presence sensor such as a motion sensor. This would be the act of implicit Human Computer Interaction (iHCI). The system would be properly integrated through the combination of hardwiring, as well as wireless interconnectivity. Wireless connectivity is proposed to be implemented through an appropriate set of media access control and physical layer specifics, such as IEEE 802.11[4]. The appliance we consider here is an energy efficient LED bulb.

The entire system's support framework can be thought of as everything but the mobile webpage and the actual appliance being controlled.

To operate the system, the user would use a combination of the options for iHCI and eHCI which are provided. However, the ultimate goal of the proposed project is to make the system as autonomous as possible, able to predict system "trigger" events using a decision value stored in a structured database, using a probability engine. This engine would calculate probabilities based on the training data set it has access to, as well as a dynamically updated probability distribution. The attributes whose probabilities contribute to the ultimate trigger decision the system makes (whether or not to trigger), would include the quantification of common, or even uncommon, daily activities which take place in a household. However, to limit the scope of the attributes considered, due to our project only being at an early research

stage, we propose to only consider a handful of contributing attributes for the system. It can be intuitively hypothesized that increasing the number of attributes considered would increase the accuracy of prediction of the crucial trigger decision, but consideration must be taken that some attributes may conflict with others, repealing the notion that more data is always better data, in the context of our project.

Our database design will is proposed to contain two tables: the first one is the 'input table', which will consider attributes like entry number, date, day, time range, time and also label denoting whether at that instance of time the lights are turned on or off. The second table will be the table considering the results of the probability engine, and also a final trigger decision based on how those probabilities relate to a preset threshold. The prior probability will simply by calculated by considering the duration for which appliance was on and subtracting it from one will give the probability of the second class then we can compute the likelihood that is the actual probability whether at that at this instance of time the light should be on or off.

4. THE BAYESIAN CLASSIFIER

Heterogeneous systems in the smart home consists of the building automation system , energy management system, fire alarm system, digital surveillance system and other network based systems. Due to the involvement of many entities and the environment being partially observable, the static decision making algorithms cannot make decisions to an acceptable degree. This calls for a smart learning algorithm which accurately speculates the triggering of appliances.

A Bayesian classifier is a supervised machine learning technique which can be employed for predictive models where the environment is highly uncertain. In the context of our proposed framework, this learning technique uses a set of labeled data where information regarding a particular instance of time i.e. that is the discretized time range is fed to the system with a class label 'yes' or 'no' denoting for which instances of time the lights should be turned on. It is based on a probalistic model where we compute the probability to trigger the appliance based on a belief network as shown in Fig 4.1. It consists of a structural model with conditional probabilities, where probability of a particular node depends upon the probability of all the incoming nodes i.e. all the parent nodes. It is only an early example of the model we will use to visualize the environment, created using the Microsoft Belief Networks Tool (MSBNx).

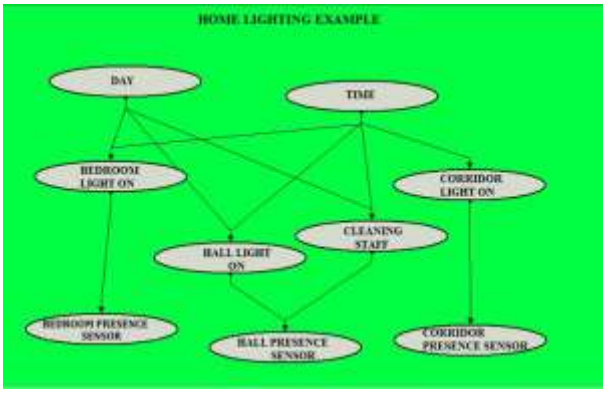


FIG 4.1: SAMPLE BELIEF NETWORK

Let us consider an example to determine in which class a set of tuples belongs based on Bayes' formula given as :

$$P(A|B) = \frac{P(A)\,P(B|A)}{P(B)}$$

Here A denotes the class which can be {yes, no}

B denotes the tuple space {B1, B2… Bn}

P (A/B) is computed by our belief network for all possible values class of A, and the maximum calculated value will be used for prediction via the function:

$$C = \operatorname{argmax}\left(P\left(\frac{A}{B}\right)\right)$$

5. CONCLUSION AND FUTURE WORK

Home automation is a technology which is still in its infancy. Here we have provided some insight into the energy conservation and automation potential of an intelligent interconnected smart home via explicit Human Computer Interaction (eHCI) i.e. mobile application and web applications, as well as implicit Human Computer Interaction (iHCI) implemented through motion sensors. Both of these methods will present different pattern in home usage patterns, which can then be modeled to predict and optimize energy needs of a living space. There are many use cases for which we require innovative solutions to use home automation systems, such as whether a user will only conduct his or her activities in a living space for a short period of time or a long period of time. Depending on what the system can "observe" about a user's habits, the system can be programmed to take appropriate actions. The problem lies in being able to efficiently give true contextual awareness to a system. Consider a case where someone is sleeping in the room with lights turned off and another person enters the room [6]. An unintelligent system will act the way it was meant to, triggering the light, which is obviously a less than desired action. The true problem is giving the system not just awareness, but also the trait of precognition, to a certain degree. Overcoming this problem could lead to massive gains

in the automation potential of a target environment, which could translate to an energy efficient deployment scenario.

The future scope of this project is also quite promising, as the basis of the project (the triggering mechanism) can easily be scaled to accommodate the needs of any household appliance, as well as any machine in a general sense after fine tuning the belief network which powers the trigger logic. Subsequently, if the system is implemented in a wide manner, such as with the backup of a cloud computing environment, it would be possible to expedite the training period of the system, by performing cluster analysis of people with similar patterns of living, and gain a large, more accurate, and quicker idea of how and when the system should trigger. A cloud supported framework would also let the user of the system be able to remotely monitor and operate the system, via the internet. Live energy statistics could be made available to the user to give the user information to act upon, and a user could fine tune the system further as he or she deems fit.

6. REFERENCES

[1] Smart Home Automation System Using Wi-Fi Low Power Devices Silviu Folea Automation Department Technical University of Cluj-Napoca Cluj-Napoca, Romania Email: silviu.folea@aut.utcluj.ro

[2] Machine Learning in Wireless Sensor Networks: Algorithms, Strategies, and Applications Mohammad Abu Alsheikh1,2, Shaowei Lin 2 , Dusit Niyato 1 and Hwee-Pink Tan 2 1School of Computer Engineering, Nanyang Technological University, Singapore 639798 2Sense and Sense-abilities Programme, Institute for Infocomm Research, Singapore 138632

[3] AFRL-IF-RS-TR-2003-25 In-House Final Technical Report February 2003 Using Bayesian Networks and Decision Theory to Model Physical Security. Nancy A. Roberts

[4] Home Automation Analysis of Current Sandeep Panigrahy, Saurabh Wahile Undergraduate Student, India, Undergraduate Student, India, International Journal of Advances in Computer Science and Technology (IJACST)

[5] Home Automation: Research Issues Dr Leslie Haddon School of Cultural and Community Studies Arts B University of Sussex Falmer, Brighton BN1 9QN UK Paper presented at the second EMTEL Workshop: The European Telecom User. November 10-11th, 1995, Amsterdam, The Netherlands

[6] http://jenson.org/easyhard/

[7] http://inhabitat.com/infographic-the-benefits-of-home-automation-systems/

[8] https://www.raspberrypi.org/

International Journal of Computer Applications Technology and Research
Volume 5– Issue 5, 245 - 248, 2016, ISSN:- 2319–8656

Remote Monitoring and Alert System for Temperature Sensitive Products

Khalid Ammar
Department of Computer Engineering
Ajman University of Science and Technology
Ajman, UAE

Abdelrahman Gamal
Department of Computer Engineering
Ajman University of Science and Technology
Ajman, UAE

Abstract: Suppliers of temperature sensitive products are required by law to store and supply these products according to predefined temperature safe range [1], [2]. A suitable system that responds to the needs of food quality assurance inspectors have been developed and tested. The system is able to continuously monitor temperature sensitive products in food-chain supply such as production, processing facilities, transportation, distribution centers, stores, restaurants, etc. The system, continuously monitors the temperature remotely and report violations in real time to a server, minimizing the need of sending food inspectors regularly to the field.

Keywords: Remote sensing, Remote Monitoring, Temperature sensing, Cold storage, GSM Module, Microcontroller, Temperature Sensitive product.

1. INTRODUCTION

All temperature sensitive products (such as dairy products, meat, fish, etc.) require storage within a certain temperature safe ranges. Failure in maintaining the right temperature throughout the food chain can lead to medical significant problems. The proposed system provides a technology that adds an extra level of protection in monitoring and tracking this temperature remotely, the system continuously report any violation in real time. The system can be used in food-chain facilities therefore a quick action can be taken by food inspectors and food managers to preserve the safety and the quality of such products.

2. THE IMPORTANT FEATURES OF THE SYSTEM

- The system provides a reliability advantage compared to manual measurements, which rely on human intervention.
- The system minimizes the need of inspectors to keep visiting sites to check the efficiency of the cooling systems.
- The system provides 24 hour continuous monitoring without any interruption.
- The system flags temperature violations in real time.
- The system uses the existing mobile network.
- The system track the refrigerators history on PC.
- The system is easy to install, use and maintain; only Sensors [5], [6], Microcontroller [3], a SIM card, and a GSM [4] connection are needed at the field side. At the server side only a PC, Microcontroller, a SIM card, and a GSM connection are needed.
- The system is scalable and can handle any size of food-chain facilities.

3. THE SYSTEM DESCRIPTION

The system has been designed and implemented as a result of the integration of various hardware and software technologies. Figure. 1 illustrates the overall system functional block diagram. The system is divided into two main parts, a field side unit and a server side unit. The field side unit collects the temperature data and sends temperature violations messages to the server side. The server side unit receives the violation messages and stores them in the data base. We utilized the Adriano Mega 2560 board which comes with a high performance internal microcontroller ATmega2560 [3]. The board, along with DHT11 sensor and GSM shields, constitute our field side unit which communicates with the server side unit.

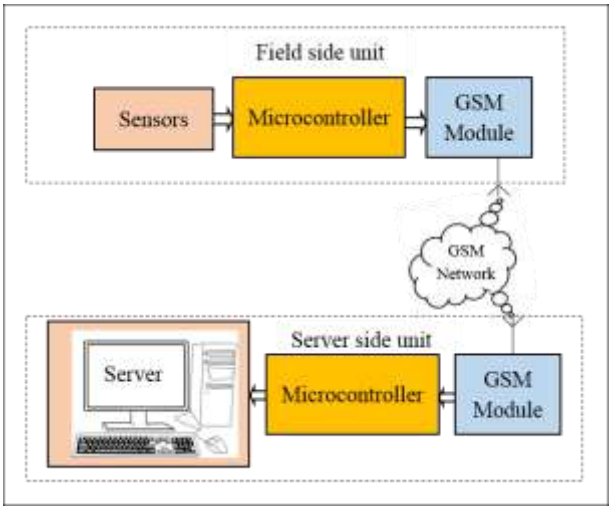


Figure. 1 The system functional block diagram.

3.1 The field side unit

The field side unit of Figure. 2 is composed of a field side hardware module and a field side software module.

3.1.1 The field side hardware module

The field side hardware module consists of three sub modules, sensor module, microcontroller module and GSM shield module.

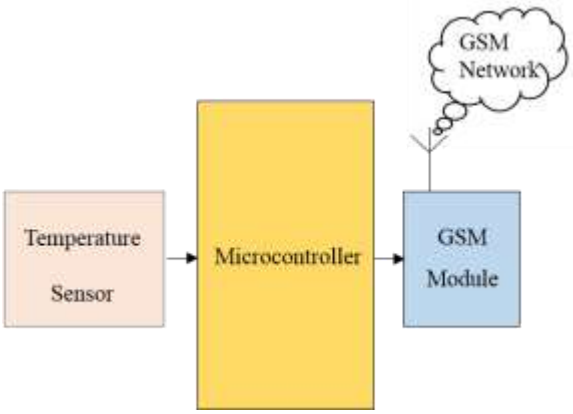


Figure. 2 Field side hardware Circuit diagram.

The microcontroller module with its embedded software collects the temperature data periodically from the sensors, compare it with the temperature set points provided by the user. If the temperature is not on the desired range, the microcontroller sends temperature violations messages to the server side via GSM/GPRS modem in real time.

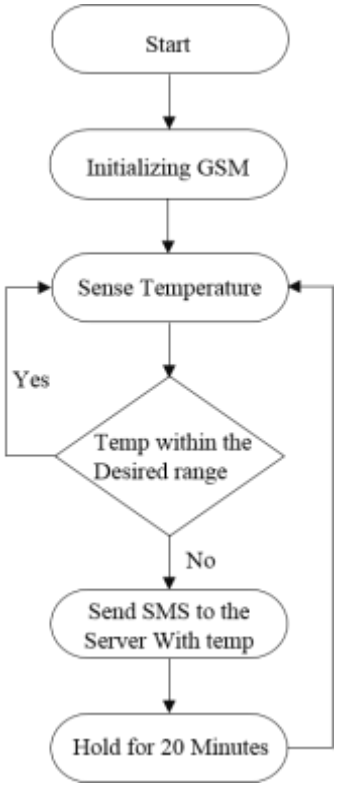


Figure. 3 The field side software flowchart

3.1.2 The field side software module

The field side software module is outlined in Figure. 3. The module initialize the GSM shield and connect the field side module to the network, after the module is connected to the network, the software module reads the temperature from the sensors. Each reading will be compared to a predefined desired temperature range. If the temperature is within the allowed range, it will take another reading, if the temperature is not within the allowed range, it will send a temperature violation message to the server including the temperature and the site ID. Figure. 4 shows an example of a violation message sent from the field side unit to the server side unit. The 32 digits identifier is a number that is stored in the microcontroller and also in the server, and it's used to identify that the message is received from authentic sender, if the same identifier does not exists in the message , the message will be discarded.

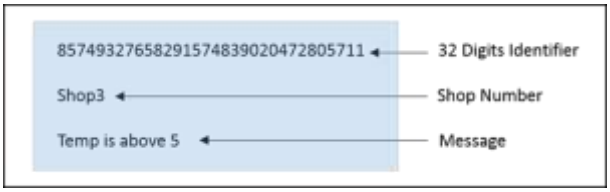


Figure. 4 Violation message example.

3.2 The server side unit

The server side unit Figure. 5 is composed of a server side hardware module and a server side software module.

3.2.1 The server side hardware module

The server side hardware module consists of three sub modules, the host computer module, the microcontroller module and the GSM shield module. The host computer acts as a server to run the server application program and to host the microcontroller and the GSM module.

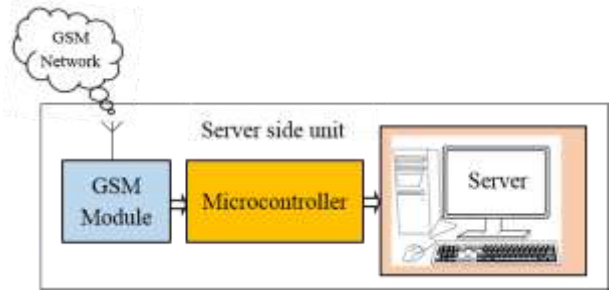


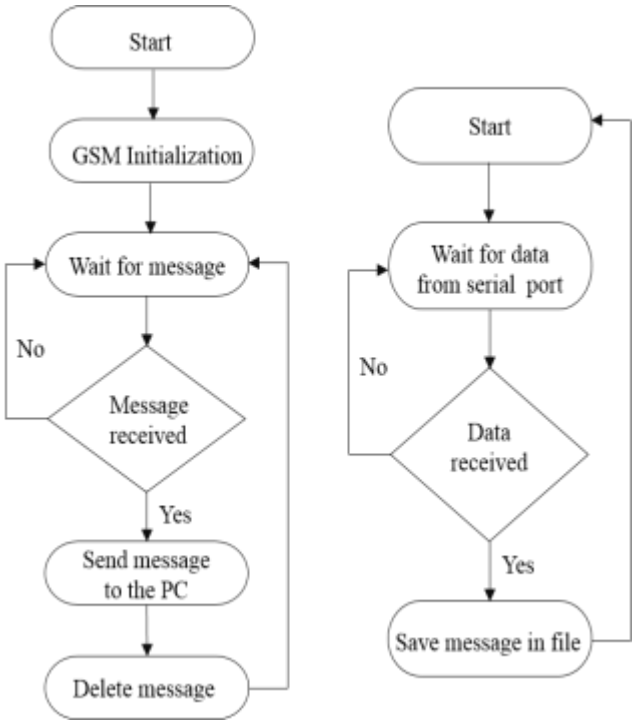
Figure. 5 The functional diagram of the server side unit.

3.2.2 The server side software module.

The server side software module consists of the microcontroller application program, the server application program, the database and the user interface.

3.2.2.1 The microcontroller and the server application program.

The overall flow chart of the microcontroller application program and the server application program are shown in Figure. 6. Both codes execute simultaneously, the microcontroller code initializes the GSM shield and connect it to the network; after the GSM is initialized the microcontroller will be ready to accept the violation messages from the field side units, if a message received, the microcontroller sends the message to the host computer. As soon as the message is received, the host computer module software will save it on the data base



(Microcontroller software nodule) (Host-Computer software nodule)

Figure. 6 The overall server side software module.

3.2.2.2 The database.

All of the temperature violations records are stored on a database for inspection. When the application initializes for the first time, a database will be created with a table called Shops-Table, this table saves the information of all shops. **Table 1.** shows a Shop-table, after adding a shop, another table will be created for this shop, to save the received violations and information. **Table 2.** shows a Violation table for specific shop, in this case shop number two.

Table 2. Shops table.

ID	SHOPID	SHOPNAME	ADDRESS	SIMNUMBER	DATEOFCRATION	CREDIT
1003	Shop1	Carrefour	Dubai	507373876	03/06/15	49.82
1004	Shop2	Safeer Hypermarket	Sharjah	508448556	03/06/15	50.0
1005	Shop3	Spinneys	Ajman	502020645	04/06/15	67.75

Table 2. Violations table for specific Shop

MESSAGENO	SHOPID	MESSAGE	DATE_	TIME_
1	Shop2	Temperature is above 5	2015/06/04	17:47:01
2	Shop2	Temp is below 0	2015/06/05	22:57:47

Figure. 7 shows the main functions that interacts with the database. The Show function is used to retrieve information from the database, and the manage function is used to add and update the information into the database.

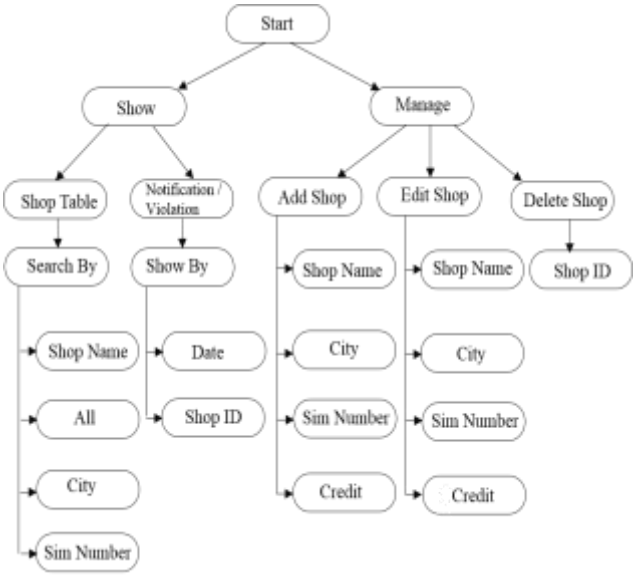


Figure. 7 Flowchart of the main functional that interacts with the database.

3.2.2.3 The user interface

A Graphical User Interface (GUI) have been developed to interact with the system by adding a site, editing a site, deleting a site and to locate the violation sites. Figure. 8 shows a sample of the GUI screenshots.

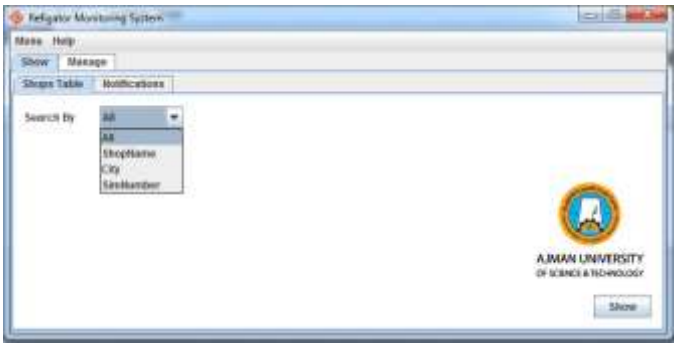


Figure. 8 Samples of GUI screenshots.

4. CONCLUSIONS

The cold storage monitoring and alert system is an indispensable tool for food managers and food inspectors. A cold storage monitoring and alert system has been fully designed and developed in house. The system provide an extra level of protection in monitoring and tracking temperature violations for temperature sensitive products remotely at any time. The system has been tested successfully against different temperatures violation setups. The system hardware and the application software's worked in harmony by taking readings from temperature sensors and sending violations to the server. The server application makes it easy for the user to retrieve and view those violations.

5. REFERENCES

[1] Food Safety Guidelines for Onsite Feeding Locations, Food Shelves and Food Banks. http://www.health.state.mn.us/divs/eh/food/fs/foodbanksafety.pdf

[2] Storing Food at the Proper Temperature Fact Sheet. https://dmna.ny.gov/foodservice/docs/toolbox/storing_food.pdf

[3] Arduino Mega 2560. http://arduino.cc/en/Main/arduinoBoardMega2560

[4] Arduino GSM Shield. http://arduino.cc/en/Main/ArduinoGSMShield

[5] DHT11 Temperature and Humidity Sensor. http://www.micro4you.com/files/sensor/DHT11.pdf

[6] DHT11 Sensor Connection with Arduino. http://playground.arduino.cc/main/DHT11Lib

International Journal of Computer Applications Technology and Research
Volume 5– Issue 5, 249 - 252, 2016, ISSN:- 2319–8656

Avoiding Man in the Middle Attack Based on ARP Spoofing in the LAN

Peta Kaleemsha
Sree Vidyanikethan Engineering College
Tirupathi, India

Pulluru Likhitha
Sree Vidyanikethan Engineering College
Tirupathi, India

Abstract: As technology is running on its wheels, networking has turned into one of our basic aspects. In this world along with networking inimical vulnerabilities are also advancing in a drastic manner, resulting in perilous security threats. This calls for the great need of network security. ARP spoofing is one of the most common MITM attacks in the LAN. This attack can show critical implications for internet users especially in stealing sensitive information's such as passwords. Beyond this it can facilitate other attacks like denial of service(DOS), session hijacking etc..,. In this paper we are proposing a new method by encrypting MAC address to shield from ARP cache poisoning.

Keywords: ARP (Address Resolution Protocol), MAC (Media access control), IP (Internet protocol), MITM, Session Hijacking, DoS, cryptography.

1. INTRODUCTION

Protecting our online data is never going to be a cake walk, especially now-a-days where attackers are regularly contriving some new techniques to sneak data. MITM attack is a serious threat to the internet users and it is far-reaching to detect as the third person, an attacker secretly places himself between two victims and he is capable of attempting any passive or active attacks. There are many techniques available today to evade MITM attack some of them are monitoring ARP traffic, making ARP cache static and highly secured data encryption. But secured data encryption cannot create a path free of ARP spoofing and it is time taking too. Instead of encrypting data we are proposing a new way of secure communication by encrypting MAC address using cryptography.

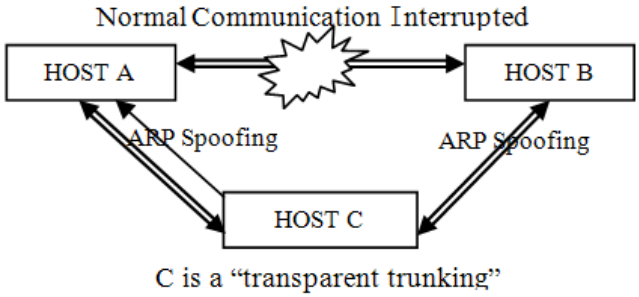
2. BASICS

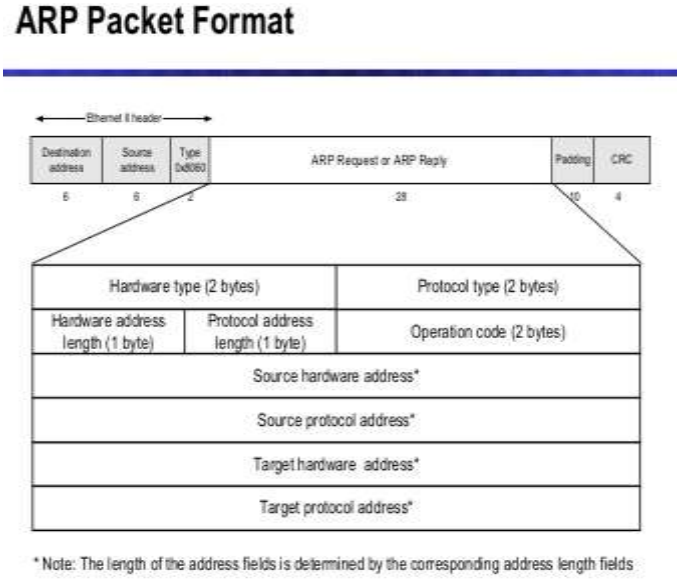
2.1 ARP communication

ARP protocol operates by broadcasting messages across a network to determine layer two MAC address of the host using predetermined IP address. Then the host with destined IP address replies back his MAC address. Thus ARP messages provide a link between IP and MAC addresses. ARP will work across a bridge. Bridges will propagate the ARP broadcasts and bridge the replies. A router will block ARP packets. This means an ARP monitor is required on each separate network segment to detect new ethernet addresses. Network switches will pass the ARP traffic because it is broadcast traffic. This means that any computer on the network can see the ARP broadcast traffic. The reply packets are usually returned directly to the requesting computer, and network switches will block this unless it is addressed to your station[1] [2].

2.2 ARP cache poisoning

It is one of the oldest forms of MITM attacks where attacker on the same subnet can eavesdrop on network traffic between the victims. It takes the advantage of insecurity indulged in ARP protocol. The malicious attacker sends a falsified ARP message and links his MAC with attacker IP address, then the attacker takes the role of "MITM", any traffic intended for that right source is send through attacker system. As it occurs at lower level the end user is oblivious to its occurrence[1]. Let A, B are two legitimate hosts communicating with each other and C be an attacker who tries to steal the information acting in between them as shown in below figure.





2.3 Denial of Service (DoS):

DoS refers to an attack that overwhelms system with data-most commonly a flood of simultaneous requests send to a website, causing the server to crash or simply becomes inoperable as it struggles in responding to more requests than it can handle. As a result, legitimate users who try to access the websites control by server are unable to get the service. There are other types of DoS attacks that use different tactics, but they all have the same effect: preventing legitimate users from accessing the site[1][4].

2.4 Session hijacking

Session hijacking attacks consists of the exploitation of web sessions controlling mechanism, which is normally managed for a session token. As http communication uses many TCP connections the web server needs a method to recognize every user's connection. The most useful method depends on the token that the web server sends to the client browser after a successful client authentication. A session token normally consists of string of variable width and it can be in different ways like in the URL. The session hijacking attack compromises the session token by stealing or predicting a valid token to gain unauthorized access to the web server[1][4].

2.5 Cryptography

Cryptography is a method of information hiding and verification to provide security from unauthorized access. It includes the protocols, algorithms and strategies to securely and consistently prevent or delay unauthorized access to sensitive information and enables verifiability of every component in a communication. It is the powerful techniques available today to go against information attackers. The two techniques of cryptography are substitution and transposition[3].

2.6 Substitution

Substitution is a method of encoding by which units of plain text will be replaced with cypher text using some fixed key, the units may be single letters, pair of letters, triplets of letters or mixture of above. As the origin letters in the text are replaced. It cannot be deciphered without knowing pre used key[3][5].

Eg: Plain text: *NETWORKING*

KEY:**2**

ENCRYPTED OUTPUT: *PGVYQTMKPI*

2.7 Transposition

In cryptography, a transposition cipher is a method of encryption by which the positions held by units of plain text are shifted according to a regular system, so that the cipher text constitutes a permutation of the plaintext. That is, the order of the units is changed. Mathematically a bijective function is used on the character's positions to encrypt and an inverse function to decrypt it[3][5].

Eg: Plain text: *NETWORKSECURITY*

KEY: *HACK*

H	A	C	K
3	1	2	4
N	E	T	W
O	R	K	S
E	C	U	R
I	T	Y	

ENCRYPTED OUTPUT: *ERCTTKUYNOEIWSR*

3. PREVIOUS WORKS

3.1 Making ARP cache static

One way to protect against ARP poisoning is to change the unsecured dynamic nature of ARP cache into a static thing. It helps only when your network configuration doesn't change often and when it is pretty simple to make a list of static ARP entries. This will ensure that device will always rely on their local ARP cache rather on ARP requests and replies[2].

3.2 Monitoring ARP traffic

This involves monitoring the network traffic of host by a third party using some intruder detection software's such as **Snort** but it is feasible when only single host is considered and it can be a bit cumbersome when entire network is concerned[2].

3.3 Virtual Private Networking

A VPN is a technology which creates an encrypted connection over a less secure network. It protects your internet and lets you browse with privacy. It extends a private network across a public network. It enables a computer to send and receive data across shared or public networks as if it were directly connected to the private network The benefit of using a VPN is that it ensures the appropriate level of security to the connected systems when the underlying network infrastructure alone cannot provide it A VPN is created by establishing a virtual point-to-point connection through the use of dedicated connections, virtual tunneling protocols or traffic encryptions, such as PPTP (Point-to-point Tunneling Protocol) or Internet Protocol Security (IPSec). The most common types of VPNs are remote access VPNs and site-to-site VPNs. Using a VPN will shut down many of the places where a MITM attack might happen, but not all of them. Specifically, it will protect your traffic between your device and the VPN gateway, preventing your ISP (or most governments) from performing a MITM attack targeted toward you. However, once your traffic passes from

the VPN gateway to its eventual destination, it becomes vulnerable to a MITM attack. With a VPN, your traffic is then semi-anonymized, so it is much more difficult to target any attack toward any particular person, but an indiscriminate attack against all users of a particular website is still very possible. And it is also too expensive to setup[4][5].

3.4 Secure Shell Tunnelling

SSH tunnel transfers unencrypted traffic over an encrypted channel created using SSH protocol connection. Generally, SSH is used to securely acquire and use a remote terminal session. For setting up a tunnel the port of one machine needs to be forwarded to a port in the other machine which is at other end of the tunnel. This uses different kinds of port forwarding mechanisms namely Local port forwarding, Remote port forwarding, Dynamic port forwarding. SSH tunnel is often called as "Poor Man's VPN" as it provides some of the same features as a VPN without more complicated server setup process however an SSH tunnel doesn't offer all the benefits of a VPN. Unlike with a VPN, you must configure each application to use SSH tunnel's proxy. With a VPN we are assured that all traffic will be sent through it - but you do not have this assurance with an SSH tunnel. And it also doesn't offers high security as data is in unencrypted form[4][5].

4. PROPOSED IDEA

In our method we are trying to eliminate MITM attack especially ARP poisoning by using switch as a third party to encrypt the mac address of the host.so that the attacker cannot interfere. The technique used for this method is as follows:

Algorithm:

let Ceaser(C), Harrison(H) are two hosts and Shlomo(S) as a third party switch. consider a scenario where Ceaser wants to communicate with Harrison.

step 1: Ceaser sends to Shlomo his mac address attached with some key k1 and IP address of Harrison, the destination.

step 2: Shlomo broadcasts the destination IP address and waits for the reply.

step 3: Then Harrison replies with his MAC address along with his key k2.

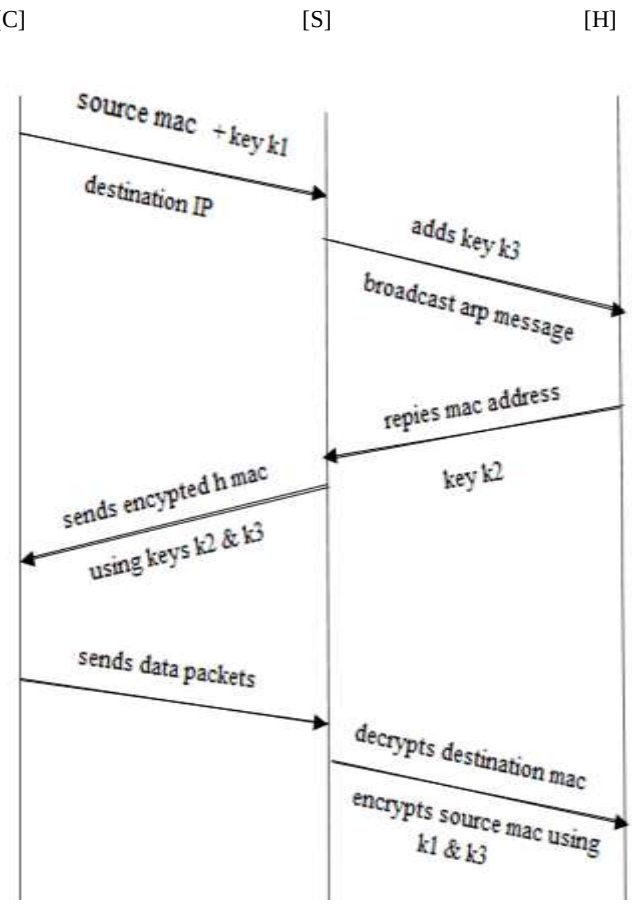
step 4: Instead of giving actual MAC address of Harrison, Shlomo encrypts it by using transposition technique with key k2 taken from him. Shlomo also adds substitution security by using his own static key k3.

step 5: Shlomo uses the same technique and encrypts the mac address of Ceaser by using transposition with key k1 and also adds substitution security with his private key k3 and gives it to Harrison.

step 6: Now the connection was established between Ceaser and Harrison without knowing their actual MAC addresses.

step 7: For replying back Shlomo decrypts the MAC addresses and goes to the right person.

step 8: For further security, amelioration we can use randomness in keys provided by hosts for different connection establishments.



5. CONCLUSION

In our technique only third party, switch knows the actual MAC addresses of the hosts in the subnet. As we are using two keys one from the host and the other from the switch, security hardens and any attacker who wants to sniff the target system's MAC address can see only the encrypted MAC. Thus there will be a great chance to get rid of ARP spoofing with our method.

6. REFERENCES

[1] Sean Whales, "*An Introduction to ARP Spoofing [EB/OL]*",http://packetstormsecurity.org/papers/protocols/, 2001.

[2] Guo Hao and Guo Tao, "*Principle of and Protection of Man-in-the-middle Attack Based on ARP Spoofing*", Journal of Information Processing Systems, Vol.5, No.3, September 2009.

[3] Whitfield Diffie and Martin Hellman, "*Multi-user cryptographic techniques*", [Diffie and Hellman, AFIPS Proceedings 45, pp109-112, June 8, 1976].

[4] Behour A. Forouzan, Sophia Chung Fegan ―" *Data Communication and Networking*", Fourth Edition 2009.

[5] William Stallings,"Cryptography *and Network Security",* Fourth Edition 2006.

International Journal of Computer Applications Technology and Research
Volume 5– Issue 5, 253 - 259, 2016, ISSN:- 2319–8656

Forest Area Estimation in Kutai Nasional Park of East Kalimantan Using Computer System Aplication based Genetic Algorithm-Support Vector Machine

Lapu Tombilayuk
Informatics Department, Sekolah Tinggi Teknologi Bontang,
Bontang East Kalimantan, Indonesia

Abstract : The paper presents design of computer system application for the forest area estimation using the combination of genetic algorithm (GA) and support vector machine SVM) methods in Kutai National park of East Borneo. The considering variables such as reboization concerning natural green, forest fire, encroachment and illegal logging activities are the basic data for our proposed design. In the development of design computer system application, the Unified Modeling Language is adopted with the use case, activity diagrams and sequence diagrams are systematically followed in order to keep the purpose of design on track. The supporting instrument for this research is the language programming of Borland Delphi 7 and MySQL database. The accuracy of area estimation result is compared with the actual data using mean absolute percentage error (MAPE).

Keywords: GA-SVM methods, UML, Borland Delphi, MySQL database, white box testing.

1. Introduction

Indonesia has continuously experienced in deforestation area where the majority of forest damaging occurs in Borneo and Sumatera. The causes of deforestation are forest mismanagement, illegal logging, forest fire and forest opening for farming and mining purposes. The effects of forest damaging can be locally, such as ecologic disaster, flood, soil avalanche and can be globally as well, such as drought and other global warming effects. The government is actually proactive to do the forest conservation program with policies, for instance the protection of primary natural and peat forests and the strict regulation of forest utilization for mining purpose. The government regulation on forest conservation is highly protected by law; in fact, deforestation seems uncontrollable. The organization of forest watch Indonesia (FWI) reported that the forest in Java will be used up in 2020 with small width area remaining in Bali and Nusa Tenggara (0.08 million h.a), Sulawesi (7.2 million h.a), Sumatera (7.72 million h.a), Borneo (21.29 million h.a) and Papua (33.45 million h.a). The rate of deforestation is about 1.51 million h.a per year in the period of 2000-2009 with the highest rate of 0.55 million h.a per year occurs in Borneo.

Forest area estimation method has gained important awareness of researchers with variety methods. The statistic methods are still dominant in this topic. With the hierarchy of Bayesian model, the forest of area can be accurately classified up to 88%. Another approach by the forest sampling method is used to estimate the forest canopy cove according to probability theory. In addition, the forest cover estimation based the importance vegetation type has been proposed using K-means method on the time series data. The research is focused on the clustering of different type of vegetation. Meanwhile, the estimation of forest canopy by comparison of the field measurement technique has been conducted by pictures result of digital camera.

The results of the previous research motivate us to find the best method to estimate the forest area based on the conditions of reboization concerning natural green, forest fire, encroachment and illegal logging activities. The parametric method by mean mathematic model and statistic, such as autoregressive integrated moving average (ARIMA) is less suitable for this case study due to the difficulty in modeling irregular and variable number of data. On the other hand, the non-parametric method with exponential technique is only superior for the short-term forecasting and the results may not be confirmed optimal. In addition, the artificial neural network is facing difficulty to provide solution with time-series data.

2. Configuration of proposed system

The time series estimation method has continuously attracted serious attention from scientific community. The method is basically part of computational intelligence utilizing historical data to solve estimation problems. This kind of technique is possible supporting fromthe advanced computational technique and information technology. One of the computational techniques based machine learning is the Support Vector Machine (SVM). This method is superior to classify the input data set from the minimum to the maximum values. The classification results are used as chromosome for the genetic algorithm (GA) operation. This combination accelerates the computational efforts and provides high accuracy estimation compared to the only GA process. In this case, the output genetic algorithm (GA) is the optimal

estimated value of forest area. It also implies that the uncorrelated data classification is avoided, as results only the important inputs are considered by the implementation of GA-SVM method. In this research, the main configuration of the estimation method for the forest area in Kutai National park, East Borneo is divided into the input database of including positive, like reboization concerning natural green and negative causes, such as forest fire, encroachment and illegal logging activities that affecting to the forest area, processing database using GA-SVM methods and the area forest estimation output. The database system is stored using MySQL software system. The genetic algorithm utilizes these data to initialize parameter and to generate the population. Later, the support vector machine method evaluates the fitness function in order obtain the data set. Then, the data set is reselected to obtain two chromosomes with the best fitness function by the genetic algorithm. By this approach, the only correlated data is used for the GA process, results in high accuracy estimation. The process and evaluation continue with crossover and mutation of new generation until it convergences at the best intent of fitness function.

The input database is taken directly from Kutai National Park office during the last 10 years (between 2003 and 2012) about data record of reboization, forest fire, encroachment and illegal logging activities.

These data is quite random and irregular due to the cause combination between the nature and human activities. In these

data, the average forest damaging area based on the negative causes is about 125.68 h.a, with the mean area of reboization is about 54.5 h.a. In 2013, the total forest area of Kutai National ark, East Borneo is about 198,629 h.a with damaging area about 711.8 h.a. Based on this reason, it is important to provide some tools to estimate the forest area for some time in the future, so that it becomes reference to measure and to accelerate the green activities inside the National Park.

The forest area estimation is performed by the implementation of combination between the genetic algorithm (GA) and support vector machine (SVM) method based on the storing data in the database system. The superiority of this method is in the capability of SVM method to divide vector space into hyperplane according to the data trend from small area to wide area classes. The algorithm combination for such estimation technique is able to provide better solution compared to other estimation techniques, such as artificial neural network because the error and generalization of the SVM method is not depending on the input vector. The complete process of GA-SVM methods is shown by the pseudo-code as follows:

1) Initialization process by the genetic algorithm to select the chromosome candidates of data stored in database.
2) Initialization of data set by running the SVM method to search all data set about reboization, forest fire, encroachment and illegal logging activities between 2003 and 2012 in the data base system.
3) Run the polynomial Kernel function to classify the non-linear data into two classes by $(\boldsymbol{XT.Xi} + 1)P$. The results are about the influence causes, measured from the most to the less value impacts.
4) The training process of all selected data set by means the influenced parameters to the forest area. The selection is aimed to obtain the hyperplane of $y(x)f(x)=1$ that separates 2 classes. The candidates of support vector are:

$$y(x)f(x) \leq +\beta+1 \text{ and } y(x)f(x) \geq 1 \quad (1)$$

where $\mathbf{Xa=Xa \cup Xb}$ and β is the arbitrarily determined value by users.

If the re-training process is conducted, then the previous training results are improved by taking only some data ($X \varepsilon Xa$) as the support vector candidates.

5) The data classification after training process is divided into $xi.w+b \geq 1$ for the 1st class and $xi.w+b \leq -1$ for the 2nd class.
6) Chromosomes selection. The best chromosome is usually selected by objective function evaluation with defined high probability. The roulette-wheel method is used in this step.
7) Fitness function evaluation by:

$$Fitness = C - f(x) \text{ or } Fitness = \frac{C}{F(x)+ \epsilon} \quad (2)$$

where C is a constant and Є is small number to avoid zero division. In this research, The fitness function = forest fire + encroachment + illegal logging – rebozation

8) Selection process with Linier Fitness Ranking (LFR).
9) Cross-Over process. One-cut point method is used to exchange the gen of the parent chromosome with cross-over probability (Pc) of 0.25.

```
Begin
  k ← 0;
  While (k < populasi) do
  R[k] ← random [0 - 1];                (3)
  If (R[k] < Pc) then
  Select Chromosome[k] as parent;
End;
```

10) Mutation process. The mutation rate is specified at 0.1.
11) Population replacement by the new generation
12) If the new generation convergences to the optimal solution, the overall process stop; otherwise the process is repeated from no. 10.

The above pseudocode is translated into language programming of Borland Delphi 7. The mean absolute percentage error (MAPE) is used to validate the accuracy in output measurement. In addition, the white box testing is used to evaluate all computer logic programming by checking the logical iteration and to assess the overall data used in the simulation.

3. Design of Computer Application for Forest Area Estimation

In this section, it will be explained about the designs information related to the unified modeling language (UML), database and user interface. The detailed information of such design is provided as follows:

3.1 *Design of Unified Modeling Language (UML)*

The unified Modeling Language (UML) is the standar design and documentation of software system with the capability of software application modelling according to the hardware configuration, operating system and it has the flexibility ain network design and language programming. The UML diagram can be built-up using use case, class, activity, sequence and statechart diagrams. With the UML design, the programmers may expect the model runs perfectly and accurately with considering the scalability, robustness and security. The use case diagram represents the expected functionality of designed system and also can be considered as the interaction mechanism between the user and the system. It consists of case login to access the system, case view menu with several menu options, case input data of such the causes that influences to forest area, case analysis by means the GA-SVM method and the case output as the estimated forest area. The input data processing may be performed by an administrator and the output data is utilized by the National Park Authority in order to anticipate the negative causes and continuing promote the green activities. The use case diagram is shown in Figure 1.

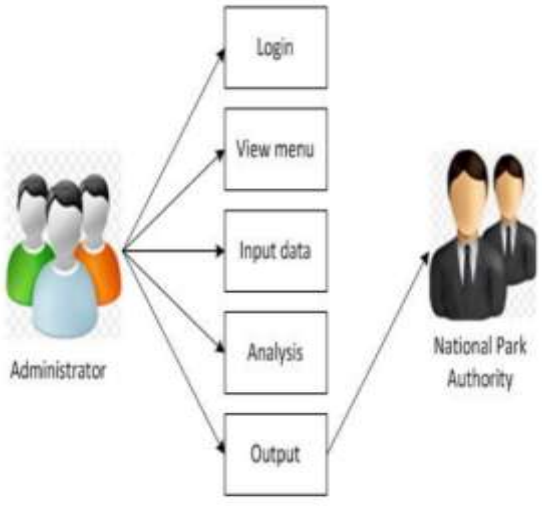


Figure 1. Use Case Diagram (4)

The class diagram represents the class of structure and its description, package and object including the connectivity within the application design. Figure 2 shows the class description of our proposed application. There are 3 data classes, i.e input data, analysis and report. The class of input data consists of causes identity, year, location, causes and width area fields. Meanwhile, the class of analysis is the fields of number, year, causes and width area. The last class is the class of report which comprises of fields of report identity and its type. The chart of class diagram is shown in Figure 2.

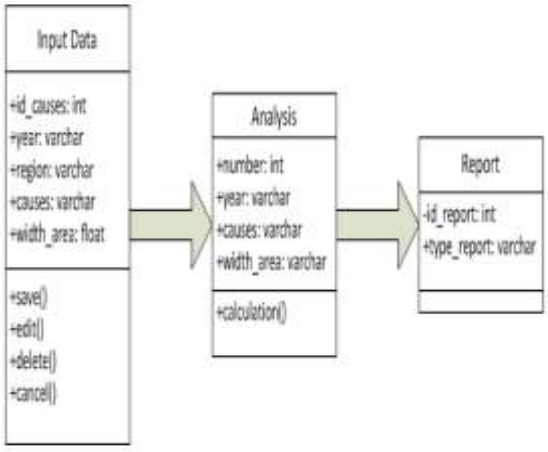


Figure 2. Class diagram

The next diagram in UML development is the activity diagram that describes the activity path inside the designed system, to how the system starts to end including the possibility decision that might be appear.
Amongst the developed diagram, the activity diagram is the special state diagram where most of the states are an action, triggered transition and internal processing by the previous states. In this research, the activity diagram explains the system process from the data inputting by an administrator, data processing using GA-SVM method and the estimation output utilization by the National park authority. The activity diagram is shown in Figure 3.

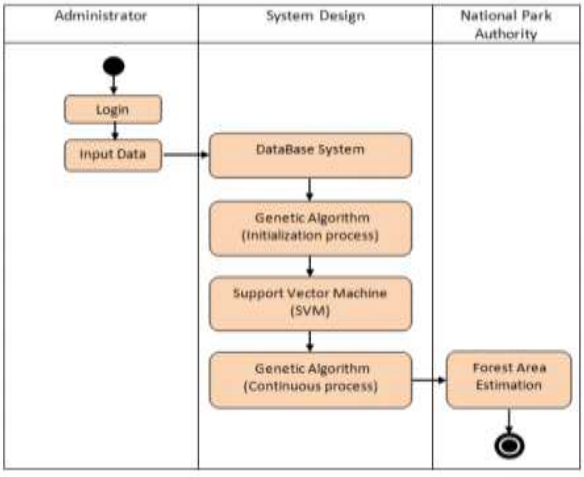


Figure 3. Actifity Diagram

The object interaction inside and surrounding the system including the user, display and so on is described by sequential diagram in the forms of messages per time value. The sequential diagram can also be used to express the scenarios or steps that should be followed according to the output events. It may start from the causes that trigger the next activity, to what kind of internal changes and typical. Each object including administrator has vertical life line. The message is depicted by arrows from an object to other objects. In the next phase design, such messages are mapped into operation or method of classes. Figure 4 explains the object behavior when the operator interacts with application system. Administrator needs login to input the causes that influence to the total forest area into the database. Later, this data is processed by the GA-SVM method to yield total forest area estimation which can be used by the National park authority.

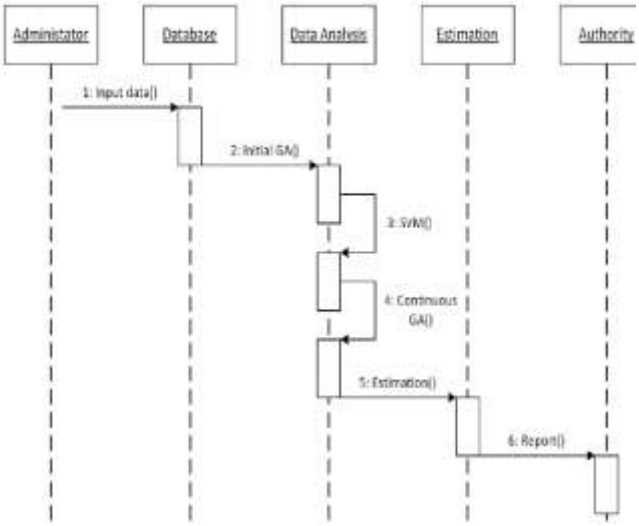


Figure 4. Sequential diagram

To support all diagrams process, it requires the state chart diagram input data that describes transition states or changes of object inside the systems. The state chart diagram is shown in Figure 5. In this figure, if the operators makes mistake during data inputting process, they have opportunity to do editing then resave the data into the database. If the cancelation of inputting data occurs, the system will terminate. The inputting data process into the database keeps going until the data is finished and saved. The remained process is just waiting the execution of GA-SVM method to yield the estimated forest area. In addition, the state chart diagram may describe the data analysis using GA-SVM method and report analysis as the outcome of algorithm.

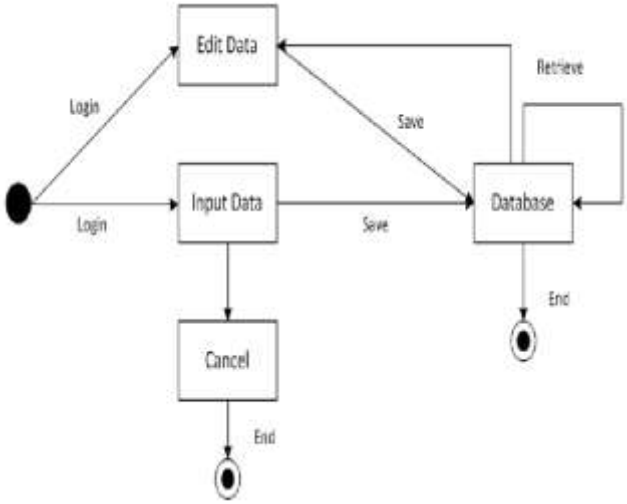


Figure 5. Statechart diagram of input and edit data

3.2 *Design of database*

The database design is used to map the conceptual model into the basis data model. In this research, the structure data file about the causes affecting the forest area located in the database is shown in Table 1. There are five field names but the most important field (primary) is the field of number. When transforming these data into terms of generation data in Genetic Algorithm called the data identity of generation, there are four field names. Again, the most important field (primary) is the field of number. The data identity of generation is shown in Table 2.

Table 1. Data identity of causes influenced to the forest area

Field name	Type	Length	Key	Declaration
Number	Int	11	Primary	The code of causes
Year	varchart	5	-	The year of events
Region	varchart	50	-	Forest region
Causes	varchart	50	-	Causes to the forest area
Width area	float	0	-	Damaged forest area

Table 2. Data identity of generation

Field name	Type	Length	Key	Declaration
Number	Int	11	Primary	The code of causes
Generation	varchart	5	-	Number of generation
Causes	varchart	50	-	Causes to the forest area
Width area	floa	0	-	Damaged forest area

In addition, all data in the database system in cooperating with software system is stored with MySQL basis management system. MySQL system is free software (open source) under the license of general public license (GPL) both the source code and executable program. The other advantage of MySQL are portability, multiuser, flexible tuning performance, high security and scalable. Therefore, design such computer application system is nowadays very convenient with maximal performance.

3.3. *Design of User interface*

The proposed user interface is the displays of the main menu, input data, initial data, generation and estimation. These displays are explained as follows. The display of main menu for the forest area estimation in Kutai National Park of East Borneo as shown in Figure 6 contains the communicative information related to input data, analysis and report. The display explains the main front of application on how the users interact to the system application simply. When the main display is active, the ser may input data by simple click. If they continue for the analysis and report, they may just precede the previous activity.

Figure 6. Display of main menu

In the inputting data process, the users may input data of year, region, and causes both positive and negative. The year data is specified from 2003 to 2012 with three definite regions where region I is called Suka Rahmat, region II is Sangatta and region III is Manamang. The last required data are the causes and their affected total area. The causes influenced to the forest area in the park are reboization, forest fire, encroachment and illegal logging activities. The page display for the input data process is shown in Figure 7.

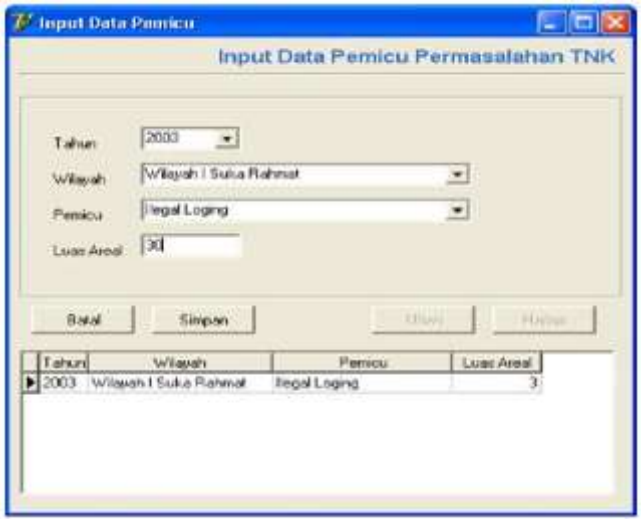


Figure 7. Interface display for input data

After the input data process is complete, the user may check their data stored in the database. The typical initial data stored in database system is shown in Figure 8. It is clearly shown that there are certain damaged forest area affected, forest fire, encroachment and illegal logging activities even though the reboization contributes positively in this matter. Total forest area in Kutai National Park, East Borneo is 198,629 h.a. However, it has been recently found that the total forest area reduced to 197,917 h.a by considering the above causes. In the current display, the users may continue the process until they obtain the estimated area in future years.

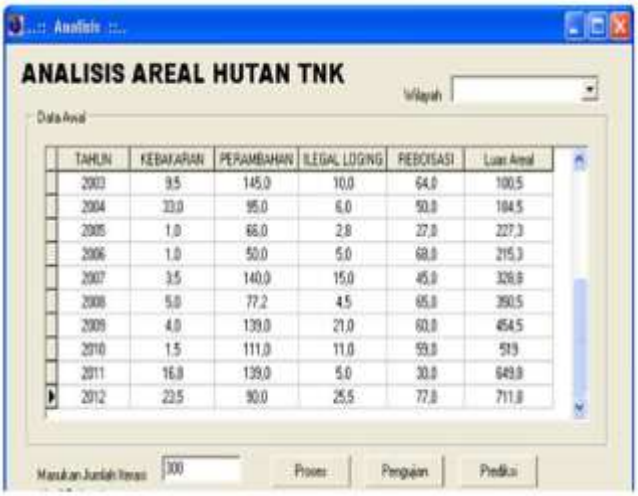


Figure 8. Initial input data display in database system

If the process continues, then the users may receive information about the estimation of forest area in the following years. In this simulation, the maximum width of forest area (the chromosome value) of the generation is the estimation output. It is due the consideration of the 10 years previous data to predict the forest area in the year after. For example in Figure 9, the chromosome value change from the initial value (data from 2003 to 2012) to other number specified that the damaged forest area in 2013 is 784.7 h.a. With the same consideration in the following generation, it will be the estimation results of the year 2014, and so on. With this application system, the estimated damaged forest area in the years of 2014, 2015 and 2016 are 905.1 h.a, 1,444.7 h.a and 2,211.2 h.a, respectively. These results are obtained with assumptions that there is no updating data from the initial data condition. The estimated results in 2016 may less than the above estimated number if the reboization activity increases and one of the negative causes can be pressed down.

Figure 9. Typical display of damage forest area estimation in 2013

If the process continues, then the users may receive information about the estimation of forest area in the following years. In this simulation, the maximum width of forest area (the chromosome value) of the generation is the estimation output. It is due the consideration of the 10 years previous data to predict the forest area in the year after. For example in Figure 10, the chromosome value change from the initial value (data from 2003 to 2012) to other number specified that the damaged forest area in 2013 is 784.7 h.a. With the same consideration in the following generation, it will be the estimation results of the year 2014, and so on. With this application system, the estimated damaged forest area in the years of 2014, 2015 and 2016 are 905.1 h.a, 1,444.7 h.a and 2,211.2 h.a, respectively. These results are obtained with assumptions that there is no updating data from the initial data condition. The estimated results in 2016 may less than the above estimated number if the reboization activity increases and one of the negative causes can be pressed down.

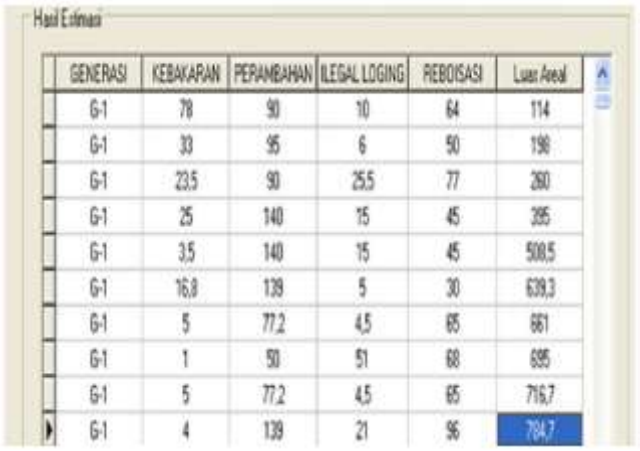


Figure 10. Typical display of damage forest area estimation in 2013

The design of application program is complete with the display of graph estimation as shown in Figure 11. In this figure, it is shown that in certain period the damage forest area decreases. For instance, there is significant reduction of forest area to about 939 h.a in 2017 and 358.4 h.a in 2020 from the previous year conditions. However, this condition is some kind of transition states because the total damaged area rises again to 1,883.9 h.a and 2,341.3 h.a in the years of 2018 and 2019, respectively. As previously mentioned that the current forest area of Kutai National Park, East Borneo is 198,629 h.a. If the prediction process continues without any changes in the input data by means there is no positive action for the green activity, the forest area is used up in 2061 because the estimated damage area in this year has reached 200,129 h.a which is far beyond the current forest area. Obtaining such number is important for the local people, park authority and global society as the reference to do positive action for the forest conservation since the Kutai National Park of East Borneo is 'the lung of world'.

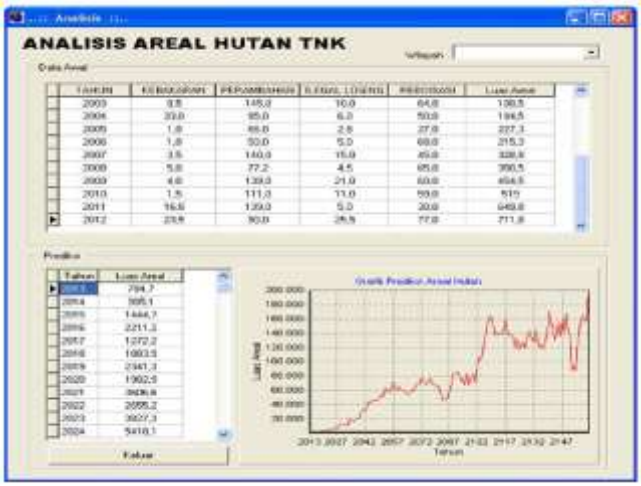


Figure 11. Typical display of estimation results with graph estimation

4. Simulation results and discussion

In the design of system application, the high accuracy estimation or prediction is the most important aspect to be considered in order to guarantee the outcomes are on the right value even the data input collection is abruptly changed. The accuracy assessment in this research is by measuring the Mean Absolute Percentage Error (MAPE) between the estimated area as the output of the designed system application and the actual area. In addition, the estimated output area is also compared with the conventional linear regression with similar MAPE performance index measurement.

The first scenario is the comparison between the estimated area and the actual area. For the easy comparison, the data of previous five years from 2003 to 2007 is arbitrarily selected as the initial input data (training data) because we have the figure actual data from 2008 to 2012 for the validation data. The result of simulation under scenario is shown in Fig. 11. In the simulation results, we have comparing data between estimated area and actual area from 2008 to 2012. These data is evaluated by Mean Absolute Percentage Error (MAPE) equation as follows:

$$\text{MAPE} = \frac{\sum_t^N \left| \frac{A_t - E_t}{A_t} \right|}{N} x100\% \quad (4)$$

where At is the actual area in the t year, Et is the estimate area in the t year and N is the period of data evaluation (N=5) in this research. The MAPE index performance calculation is summarized in Table III. For the five years period of estimation, the average of MAPE is about 2.1 % with the maximum difference of 102.8 h.a in 2008 and minimum difference of 8.8 h.a in 2011. It means that the proposed system application to estimate the forest area of Kutai National Park, East Borneo is guarantee small. Therefore, the estimated forest area until the year of 2020 may present proper results. Typical display of estimation results for the following years has been shown in Figure 12.

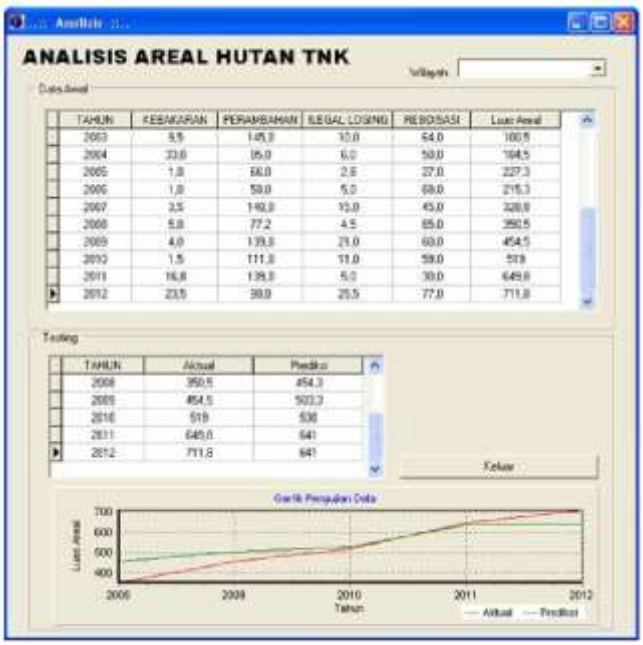


Figure 12. Simulation result for the estimated and actual area comparison

Table 3. MAPE performance index measurement

Year	Estimated area (h.a)	Actual area (h.a)	Difference (h.a)	MAPE (%)
2008	453.3	350.5	102.8	5.9
2009	503.3	454,5	48.8	2.1
2010	530.0	519.0	11.0	0.4
2011	641.0	649.8	8.8	0.3
2012	641,0	711.8	70.8	2.0

The last performance testing to our proposed system design is the White Box Testing. The test is conducted to see the module contents in order to evaluate the property of coding program. The test may cover the logical statements and their decision, iteration process within its constraints and the overall internal data structure to guarantee the validity of program. Basically, the white box testing is the path testing to allow the programmers to measure the logical complexity of procedural design and to use this measurement as the guidance to define the path set. The white box testing of this research is shown in Figure 13. There are 5 regions, denoted with R1= initial population, R2= chromosome selection, R3= crossover, R4= mutation and R5 = generation iteration. According to *the* simulator Flowgraph application, the Edge number is 14 and the Node number is 11; therefore the Cyclometric Complexity (CC) is equal to 5. Also, we obtain 5 paths from Fig. 12. which are Path 1: A – B – C – D – E – D – E – F – G – H – I – J – K; Path 2:A – B – C – D – E – F – G – F – G – H – I – J – K; Path 3: A – B – C – D – E – F – G – H – I – H – I – J – K; Path 4: A – B – C – D – E – F – G – H – I – J – C – D – E – F –G – H – I – J – K and Path 5: A – B – C – D – E – F – G – H – I – K. Because of the proposed system application has 5 regions, 5 CCs and 5 paths, the analysis application can be claimed to be properly correct.

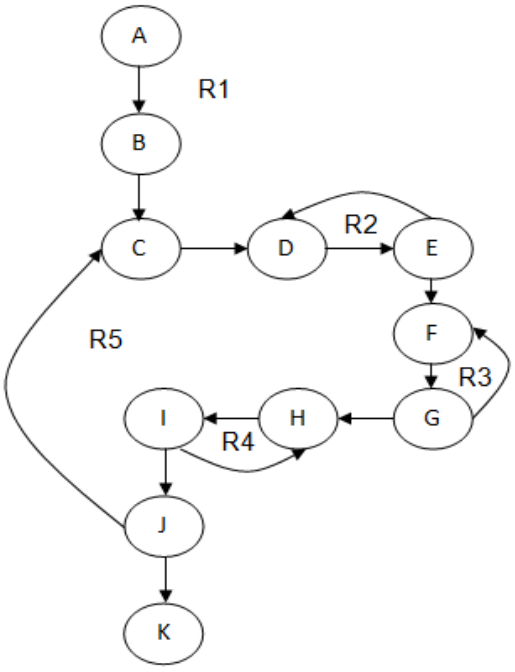


Figure 13. The flow graph in white box testing

5. Conclusion

The paper presents the design computer system application to estimate the forest area in Kutai National Park, East Borneo based the combination of genetic algorithm (GA) and support vector machine (SVM). In the development of design system application, the Unified Modeling Language is adopted with the case, activity and sequence diagrams are systematically followed in order to keep the purpose of design on track considering variables such as reboization concerning natural green, forest fire, encroachment and illegal logging as the input data in database system. The supporting instrument for this research is the language programming of Borland Delphi 7 and MySQL database. Without any actions to the initial data by means there is no significant effort to do the forest conservation program by the park authority, the forest area will be finished by the year of 2061. It is very dangerous situation to the world ecosystem because this park is 'the lung of the world'. Provision initial data information is very urgently to save our environment. The accuracy of area estimation result is compared with the actual data using mean absolute percentage error (MAPE) with the average error of 2.1%. In addition, the white box testing to calculate region, cyclometric complexity and path is implemented to confirm the correctness of the logic algorithm of design application program.

References

[1] Wirenro Sumargo, Soelthon Gussedya Nanggara, Frionny Nainggolan, Isnenti Apriani: "Potret Keadaan Hutan Indonesia Periode tahun 2000-2009", Forest Watch Indonesia, 2011, pp. 20-22.

[2] Peraturan Pemerintah Republik Indonesia Nomor 60 Tahun 2012: "Perubahan Atas Peraturan Pemerintah Nomor 10 Tahun 2010 Tentang Tata Cara Perubahan Peruntukan dan Fungsi Kawasan Hutan", 2012.

[3] Dudy Subagdja: "Nasib Hutan Kita dan Kebijakan Ekonomi Hijauh", Berita Kompasiana, 29 Maret 2013.

[4] Christina Basaria S.: "Kajian Kelestarian Tegakan Dan Produksi Kayu Jati Jangka Panjang KPH Bojonegoro Perum Perhutani Unit II Jawa Timur", Institut Pertanian Bogor, 2009.

[5] Dyah Pratiwi, et.al: "Penghitungan laju luas area hutan berbasis algoritma segmentasi warna local", Konferensi Nasional Sistem Informasi 2013, pp. 471-475.

[6] F. Deppe: "Forest Area Estimation Using Sample Survey and Landsat MSS and TM Data", Photogrammetric Engineering & Remote Sensing, Vol.6+, No.4, 1998, pp. 285-292.

[7] Dang Khoi dan Yuji Murayama: "Forecasting Areas Vulnerable to Forest Conversion in the Tam Dao National Park Region", Remote Sens. Vol. 2, 2010, pp. 1249-1272.

[8] Loghman Ghahramany, Pariz Fatehi, Hedayat Ghazanfari: "Estimation of Basal Area in West Oak Forests of Iran Using Remote Sensing Imagery", International Journal of Geosciences, Vol. 3, 2012, pp. 398-403.

[9] Oliver Diedershagen, Barbara Koch: "Automatic Estimation Of Forest Inventory Parameters Based on Lidar, Multi-Spectral and Fogis Data", Holgen Weinacker, University Freiburg, Germany, 2003, pp. 4-13.

[10] Andrew O. Finley, et.al.: "A Bayesian approach to multi-source forest area estimation", USA, Environ Ecol Stat, Vol. 15, 2008, pp.241–258.

[11] Farshad Keivan Behjou, Mahbobeh Foshat: "Using A Sampling Method for Estimation of Forest Canopy cover", International Journal of Agriculture: Research and Review. Vol., 3 (2), 2013, pp. 217-222.

[12] Anuj Karpatne, et.al.:, "Importance of Vegetation Type in Forest Cover Estimation", University of Minnesota, 2010.

[13] Lauri Korhonen, Kari T. Khorhonen, Miina Rautianen and Pauline Stenberg: "Estimation of Forest Canopy Cover: a Comparison of field measurement Techniques", Silva Fennica, Vol. 40, No. 4, 2006, pp.577-588.

[14] Thi Nguyen, Lee Gordon-Brown, Peter Wheeler, Jim Peterson: "GA-SVM Based Framework for Time Series Forecasting", the fifth International Conference on Natural Computation 2009, pp.493-498.

International Journal of Computer Applications Technology and Research
Volume 5– Issue 5, 260 - 265, 2016, ISSN:- 2319–8656

Software Defect Prediction Using Radial Basis and Probabilistic Neural Networks

Riyadh A.K. Mehdi
College of Information Technology
Ajman University of Science and Technology
Ajman, United Arab Emirates

Abstract: Defects in modules of software systems is a major problem in software development. There are a variety of data mining techniques used to predict software defects such as regression, association rules, clustering, and classification. This paper is concerned with classification based software defect prediction. This paper investigates the effectiveness of using a radial basis function neural network and a probabilistic neural network on prediction accuracy and defect prediction. The conclusions to be drawn from this work is that the neural networks used in here provide an acceptable level of accuracy but a poor defect prediction ability. Probabilistic neural networks perform consistently better with respect to the two performance measures used across all datasets. It may be advisable to use a range of software defect prediction models to complement each other rather than relying on a single technique.

Keywords: Software defect prediction; datasets; neural networks; radial basis functions; probabilistic neural networks.

1. INTRODUCTION

Defects in modules of software systems is a major problem in software development. Software failure of an executable product or non-conformance of software to its functional requirements is a software defect. A software defect is a fault, error, or failure in a software module that results in incorrect result or unexpected behavior. These defects can arise from mistakes and errors made during a software coding or in its design and few are caused by compilers producing incorrect code. Predicting faulty software modules and identifying general software areas where defects are likely to occur could help in planning, controlling and executing software development activities and save time and money [1]. In the context of software engineering, software quality refers to software functional quality and software architectural quality. Software functional quality reflects functional requirements whereas architectural quality emphasizes non-functional requirements. The objective of software product quality engineering is to achieve the required quality of the product through the definition of quality requirements and their implementation, measurement of appropriate quality attributes and evaluation of the resulting quality [2].

A software metric is a standard of a quantitative measure of the degree to which a software system or process possesses some property. Metrics are functions, while measurements are the numbers obtained by the application of metrics. Metrics allow us to gain understanding of relationships among processes and products and build models of these relationships. Software quality metrics are a subset of software metrics that focus on the quality aspects of the product, process, and project. Product metrics describe the characteristics of the product such as size, complexity, design features, performance, and quality level. Process metrics can be used to improve software development and maintenance such as the effectiveness of defect removal during development, the pattern of testing defects arrival, and the response time of the fix process. Project metrics describe the project characteristics and execution which includes the number of software developers, the staffing pattern over the life cycle of the software, cost, schedule, and productivity [3][4]. Software defect prediction refers to those models that try to predict potential software defects from test data. There exists a correlation between the software metrics and the existence of a fault in the software. A software defect prediction model consists of independent variables (software metrics) collected and measured during software development life cycle and a dependent variable (defective or non-defective software) [2].

2. LITERATURE REVIEW

Wahono [4] provides a systematic literature review of software defect prediction including research trends, datasets, methods and frameworks. There are a variety of data mining techniques used to predict software defects such as regression, association rules, clustering, and classification. This paper is concerned with classification based software defect prediction. Various classification techniques have been used such as:

- Neural Networks
- Decision trees
- Naïve Bayes
- Support Vector Machines
- Case Based Reasoning

In their work, Okutan and Yildiz [5] used a Bayesian networks to determine the set of metrics that are most important and focus on them more to predict defectiveness. They used the Bayesian networks to determine the probabilistic influential relationships among software metrics and defect proneness. In addition to the metrics used in Promise data repository, they defined two more metrics, i.e. NOD for the number of developers and LOCQ for the lack of code quality. They extract these metrics by inspecting the source code repositories of the selected Promise data repository data sets. From the model they can determine the marginal defect proneness probability of the whole software system, the set of most effective metrics, and the influential relationships among metrics and defectiveness. Their experiments on nine open source Promise data repository data sets show that response for class (RFC), lines of code (LOC), and lack of coding quality (LOCQ) are the most effective metrics whereas coupling between objects (CBO), weighted method per class (WMC), and lack of cohesion of methods (LCOM) are less effective metrics on defect proneness. Furthermore, number of children (NOC) and depth

of inheritance tree (DIT) have very limited effect and are untrustworthy. On the other hand, based on the experiments on Poi, Tomcat, and Xalan data sets, they observed that there is a positive correlation between the number of developers (NOD) and the level of defectiveness. However, they stated that further investigation involving a greater number of projects is needed to confirm their findings.

Kaur and Kaur [6] have tried to find the quality of the software product based on identifying the defects in the classes. They have done this by using different classifiers such as Naive base, Logistic regression, Instance based (Nearest-Neighbour), Bagging, J48, Decision Tree, Random Forest. This model is applied on five different open source software to find the defects of 5885classes based on object oriented metrics. Out of which they found only Bagging and J48 to be the best.

Li and others [7] described three methods for selecting a sample: random sampling with conventional machine learners, random sampling with a semi-supervised learner and active sampling with active semi-supervised learner. To facilitate the active sampling, we propose a novel active semi-supervised learning method ACoForest which is able to sample the modules that are most helpful for learning a good prediction model. Our experiments on PROMISE datasets show that the proposed methods are effective and have potential to be applied to industrial practice.

Gao and Khoshgoftarr [8], presented an approach for using feature selection and data sampling together to deal with the problems. Three scenarios are considered: 1) feature selection based on sampled data, and modeling based on original data; 2) feature selection based on sampled data, and modeling based on sampled data; and 3) feature selection based on original data, and modeling based on sampled data. Several software measurement data sets, obtained from the PROMISE repository, are used in the case study. The empirical results demonstrated that classification models built in scenario 1) result in significantly better performance than the models built in the other two scenarios. In their work, Purswani and others [9] have combined a K-means clustering based approach with a feed-forward neural network using PC1 data set from NASA MDP software projects. The performance was based on MAE and RMSE values. Results have shown that this hybrid approach is better than analytical approaches.

Artificial Neural Networks (ANN) have been used in software defect prediction. ANNs are inspired by the way biological nervous system works, such as brain processes an information. An ANN mimics models of biological system, which uses numeric and associative processing. In two aspects, it resembles the human brain. First it acquires knowledge from its environment through a learning process. Second, synaptic weights that are used to store the acquired knowledge, which is interneuron connection strength. There are three classes of neural networks, namely single layer, multilayer feed forward networks and recurrent networks. Neural networks have been shown to perform well in classification tasks. However, there are different neural networks architectures. This paper investigates the effectiveness of using a radial basis function neural network and a probabilistic neural network on prediction accuracy and defect prediction compared with other approaches [10].

3. RESEARCH METHODOLOGY

The most commonly used neural network architecture is the back propagation trained multilayered feed forward networks with sigmoidal activation function [11]. Each neuron in a MLP takes the weighted some of its input values. That is, each input value is multiplied by a coefficient, and the results are all summed together. A single MLP neuron is a simple linear classifier, but complex non-linear classifiers can be built by combining these neurons into a network.

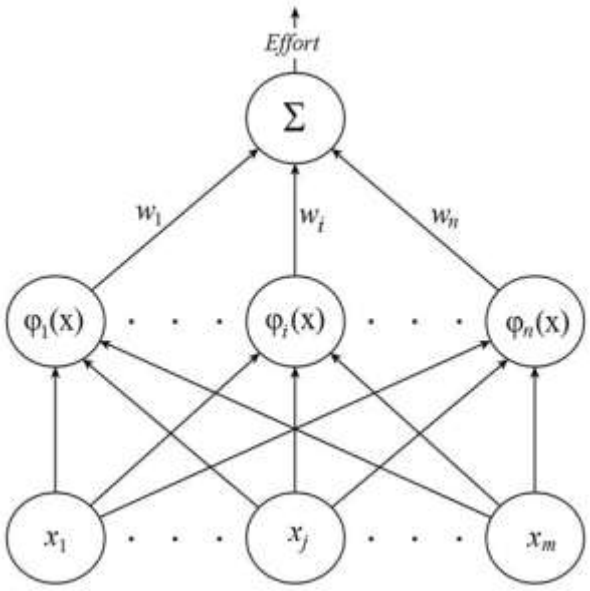


Figure 1. Radial Basis Function Neural Network

3.1 Radial Basis Function Neural Network

Another type of ANN that has been used in the literature is the radial basis functions neural network (RBFNN) [12]. A RBFNN can approximate any function which makes it suitable to model the relationship between inputs (the various cost drivers) and output (effort required). RBFNN performs classification by measuring the input's similarity to examples from the training set. Each RBFNN neuron stores a "prototype", which is just one of the examples from the training set. To classify a new input, each neuron computes the Euclidean distance between the input and its prototype. In other words, if the input more closely resembles class A prototypes than the class B prototypes, it is classified as class A. It has been show that RBFNN perform better than other types of neural networks based models [13]. In this paper we will use RBFNN to examine the effect of preprocessing datasets with PCA on the accuracy of software effort estimation models.

3.2 RBFNN Implementation

The following generic description of a RBF neural network is based on a tutorial given in [11][12]. Figure 1 describes a typical architecture of an RBFNN. It consists of an input vector, a layer of RBF neurons, and an output layer with one node per category or class of data. The input vector is the m-dimensional vector to be classified. The hidden layer consists of neurons where each one stores a "prototype" vector which is just one of the vectors from the training set. Each RBFNN neuron compares the input vector to its prototype, and outputs a value between 0 and 1 which is a measure of similarity. If the input is equal to the prototype, then the output of that RBFNN neuron will be 1. As the distance between the input and prototype grows, the response falls off exponentially towards zero.

The neuron's response value is also called its "activation" value. The prototype vector is also often called the neuron's

"center", since it's the value at the center of the bell curve. The output of the network consists of a set of nodes, one per category to be classified or a value to be computed. Each output node computes a sort of score for the associated category. Typically, a classification decision is made by assigning the input to the category with the highest score. The score is computed by taking a weighted sum of the activation values from every RBF neuron. Because each output node is computing the score for a different category, every output node has its own set of weights. The output node will typically give a positive weight to the RBF neurons that belong to its category, and a negative weight to the others.

Each RBFNN neuron computes a measure of the similarity between the input and its prototype vector (taken from the training set). Input vectors which are more similar to the prototype return a result closer to 1. There are different possible choices of similarity functions, but the most popular is based on the Gaussian function. Equation 1 describe the formula for a Gaussian with a one-dimensional input.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{--------- (1)}$$

Where x is the input, μ is the mean, and σ is the standard deviation. This produces the familiar bell curve shown below in Figure 2, which is centered at the mean, μ. In the Gaussian distribution, μ refers to the mean of the distribution. Here, it is the prototype vector which is at the center of the bell curve.

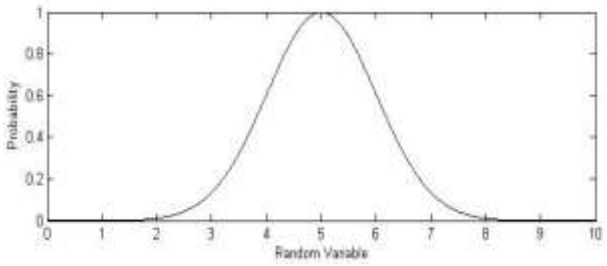


Figure 2. A Gaussian function with μ=5, and σ=1.

For the activation function, $\varphi(\ )$, the standard deviation, σ, is not important and the following two simplifying modifications can be made. The first modification is to remove the outer coefficient, 1 / (σ * sqrt (2 * π)). This term normally controls the height of the Gaussian curve. Here, though, it is redundant with the weights applied by the output nodes. During training, the output nodes will *learn* the correct coefficient or "weight" to apply to the neuron's response. The second change is to replace the inner coefficient, 1 / (2 * σ^2), with a single parameter β which controls the width of the bell curve. Again, in this context, the value of σ is not in itself important and what is needed is some coefficient that controls the width of the bell curve. Thus, after making these two modifications, the RBFNN neuron activation function can be written as in equation 2 [12]:

$$\varphi(x) = e^{-\beta\|x-\mu\|^2} \quad \text{--------- (2)}$$

There is also a slight change in notation here when we apply the equation to n-dimensional vectors. The double bar notation in the activation equation indicates that we are taking the Euclidean distance between x and μ, and squaring the result. For a 1-dimensional Gaussian, this simplifies to just $(x - \mu)^2$.

It is important to note that the underlying metric here for evaluating the similarity between an input vector and a prototype is the Euclidean distance between the two vectors. Also, each RBF neuron will produce its largest response when the input is equal to the prototype vector. This allows to take it as a measure of similarity, and sum the results from all of the RBFNN neurons. As we move out from the prototype vector, the response falls off exponentially. Every RBF neuron in the network will have some influence over the classification decision. The exponential fall off of the activation function, however, means that the neurons whose prototypes are far from the input vector will actually contribute very little to the result [12].

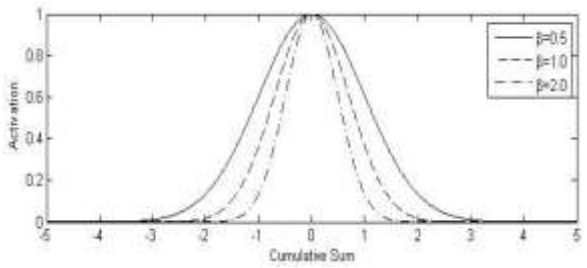


Figure 3. Activation Function for different values of beta.

3.3 Probabilistic Neural Networks

Probabilistic neural networks (PNN) can be used for classification problems. The following description of probabilistic neural networks is taken from the Neural Networks Toolbox documentation [11]. When an input is presented, the first layer computes distances from the input vector to the training input vectors, and produces a vector whose elements indicate how close the input is to a training input. The second layer sums these contributions for each class of inputs to produce as its net output a vector of probabilities. Finally, a *compete* transfer function on the output of the second layer picks the maximum of these probabilities, and produces a one for that class and a 0 for the other classes. The architecture for this system is shown in Figure 4.

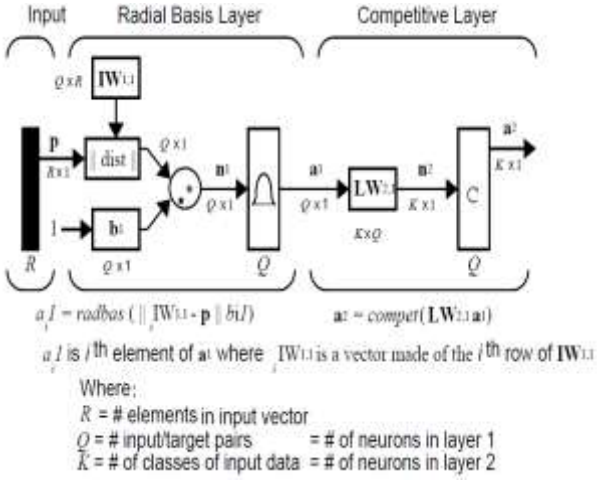


Figure 4. Probabilistic Neural Network Architecture.

It is assumed that there are Q input vector/target vector pairs. Each target vector has K elements. One of these element is one and the rest is zero. Thus, each input vector is associated with one of K classes. The first layer input weights, IW1,1 are set to

the transpose of the matrix formed from the Q training pairs, **P**'. When an input is presented the ||dist|| box produces a vector whose elements indicate how close the input is to the vectors of the training set. These elements are multiplied, element by element, by the bias and sent the *radbas* transfer function. An input vector close to a training vector will be represented by a number close to one in the output vector **a**1. If an input is close to several training vectors of a single class, it will be represented by several elements of **a**1 that are close to one. The second layer weights, LW1,2, are set to the matrix **T** of target vectors. Each vector has a one only in the row associated with that particular class of input, and zeros elsewhere. The multiplication **Ta**1 sums the elements of **a**1 due to each of the K input classes. Finally, the second layer transfer function, compete, produces a one corresponding to the largest element of **n**2, and zeros elsewhere. Thus, the network has classified the input vector into a specific one of K classes because that class had the maximum probability of being correct.

3.4 Software Defect Prediction Datasets

The following NASA software defect prediction datasets available publicly from the PROMISE repository are used in this research:

- KC1: a C++ system implementing storage management for receiving and processing ground data.
- KC2: same as KC1 but with different personnel.
- CM1: is a NASA spacecraft instrument written in "C".
- JM1: A real-time predictive ground system written in "C".
- PC1: is s flight software for earth orbiting satellite written in "C".

Defect detectors are assessed according based on the following confusion matrix:

		Modules actually has defects	
		No	Yes
Classifier predicts no defects	No	*a*	*b*
Classifier predicts some defects	Yes	*c*	*d*

Measures used based on the above confusion matrix are:

- *Acc*, Accuracy = $(a + d) / (a + b + c + d)$
- *PD*, Probability of detection $= d / (b + d)$

4. IMPLEMENTATION and RESULTS

MATLAB neural network toolbox [11] was used to implement the RBFNN and the PNN. For RBFNN, the *newrb* function is used:

Ne t= newrb (*P, T, goal, spread, MN, DF*), where

- *P* is a *m*×*n* matrix of input vector, *m* is the number of cost drivers for each project, and *n* is the number of projects used in the training phase.
- *T* is 1×*n* vector of actual efforts for each project used in the training phase.
- *Goal: mean squared errors*, 0.0001 is used.
- *Spread*: a value between 1.0 and 3.0 is used.
- *MN*: maximum number of neurons (default *n*)
- *DF*: number of neurons to add between displays, 2 is used.

For the PNN, the *newppnn* function is used:

Ne t= newpnn (*P, T, Spread*), where

- *P* is a *m*×*n* matrix of input vector, *m* is the number of attributes, and *n* is the number of input vectors used in the training phase.
- *T* is 1×*n* vector of target vectors used in the training phase.
- *Spread:* spread of radial basis functions (default = 0.1). If spread is near zero, the network acts as a nearest neighbor classifier. As spread becomes larger, the designed network takes into account several nearby design vectors.

Table 1 shows the results obtained from applying a radial basis and a probabilistic neural networks to the PROMISE datasets.

Table 1. Performance measures of applying RBFNN and PNN to the PROMISE datasets.

	RBFNN		PNN	
Dataset	*Acc*	*PD*	*Acc*	*PD*
KC1	%77.0	%44	%83.1	%29.7
KC2	%78.0	%50	%83.4	%50.0
CM1	%82.5	%20	%87.3	%33.3
JM1	%77.0	%31	%77.5	%30.0
PC1	%89.2	%32	%91.6	%40.0

Figure 5 shows the accuracy and prediction power of radial basis function neural net when applied to the PROMISE datasets. It can been seen that accuracy obtained is very close for the various datasets except for the PC1 dataset. As these data sets use the same software attributes and collected by the same processes, the difference in performance may be attributed to the nature of the PC1 project. The performance with respect to prediction of defects is not encouraging. This issue need to be investigated further.

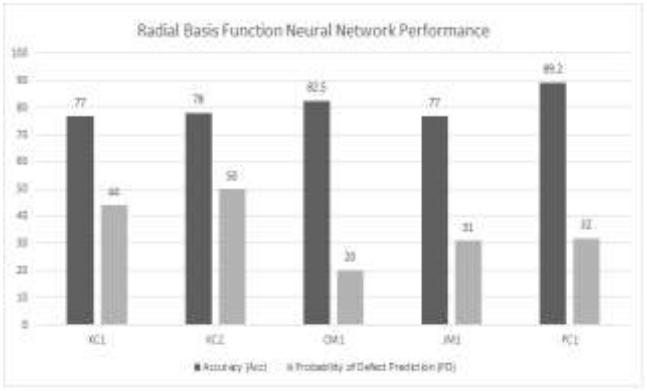


Figure 5. Accuracy and prediction power of RBF neural network to the PROMISE datasets.

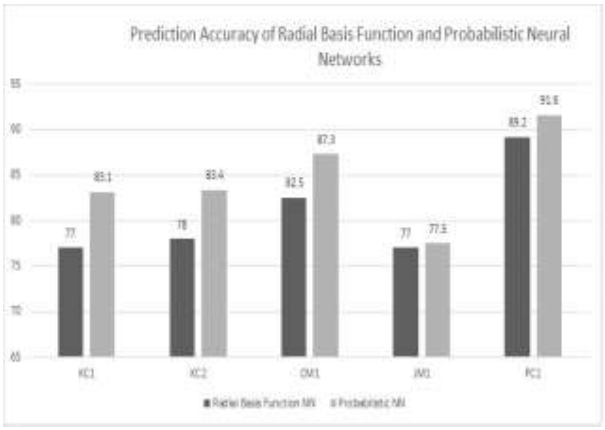


Figure 6. Actual, estimated, and PCA estimated efforts for the Cocomo81 data set.

The accuracy and defect prediction ability of probabilistic neural networks are depicted in Figure 6. Accuracy obtained is slightly better than that obtained from RBFNN. However, prediction ability is poor. However, it can be seen from Figures 7 and 8 that probabilistic neural nets performed consistently better that radial basis function neural networks with respect to accuracy and defect prediction ability.

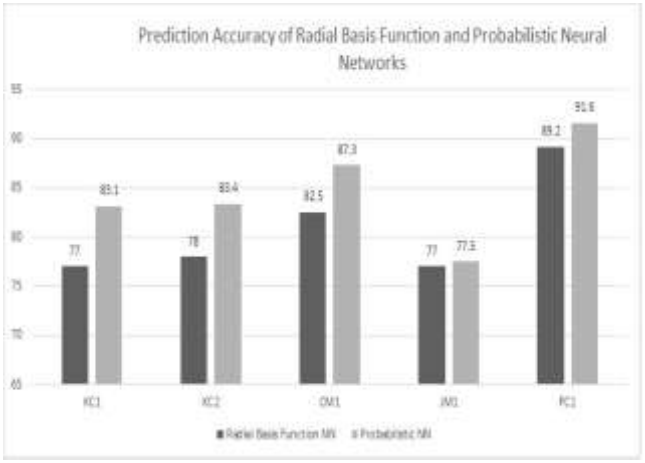


Figure 7. Actual, estimated, and PCA estimated efforts for the Maxwell data set.

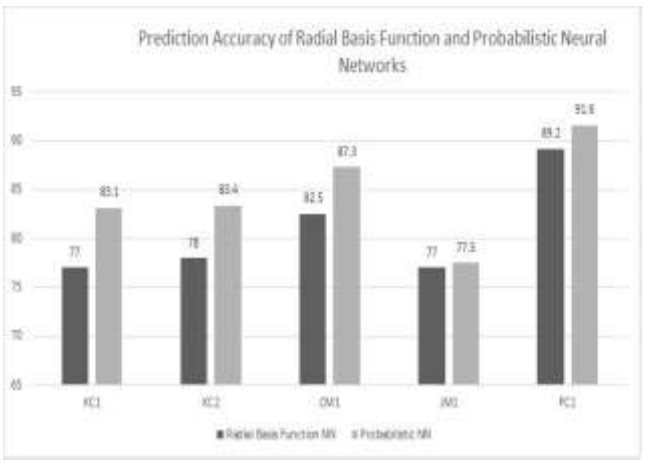


Figure 8. Actual, estimated, and PCA estimated efforts for the China data set.

5. CONCLUSIONS

The conclusions to be drawn from this work is that the neural networks used in here provide an acceptable level of accuracy but a poor defect prediction ability. Probabilistic neural networks perform consistently better with respect to the two performance measures used across all datasets. It may be advisable to use a range of software defect prediction models to complement each other rather than relying on a single technique. Further investigation of the use of neural network approached in software defect prediction is necessary to reach a solid conclusion.

6. ACKNOWLEDGMENTS

I would like to thank Ajman University of Science and Technology for providing the time and resources to conduct this research.

7. REFERENCES

[1]. K. Gupta, S. Kang, "Fuzzy Clustering Based approach for Prediction of Level of Severity of Faults in Software Systems," *International Journal of Computer and Electrical Engineering*, vol. 3, no. 6, 2011.

[2]. M. Prasad, L. Florence, and A. Arya, "A Study on Software Metrics Based Software Defect Prediction using Data Mining and Machine Learning Techniques," *International Journal of Database Theory and Application*, vol. 8, no. 3, 2015.

[3]. L. Madeyski, M. Jureczko, "Which process metrics can significantly improve defect prediction models? An empirical study," *Software Quality Journal*, vol. 23, issue 3, 2015.

[4]. R. S. Wahono, "A Systematic Literature Review of Software Defect Prediction: Research Trends, Datasets, Methods and Frameworks," Journal of Software Engineering, vol. 1, no. 1, 2015.

[5]. A. Okutan, O. T. Yildiz, "Software defect prediction using Bayesian networks," *Empirical Software Engineering*, vol. 19, no. 1, 2014.

[6]. A. Kaur, I. Kaur, "Empirical Evaluation of Machine Learning Algorithms for Fault Prediction," *Lecture Notes on Software Engineering*, vol. 2, no. 2, 2014.

[7]. M. Li, H. Zhang, R. Wu, and Z. H. Zho, "Sample-based software defect prediction with active and semi-supervised learning," *Automated Software Engineering*, vol. 19, no.2, 2012.

[8]. K. Gao, T. M. Khoshgoftaar, "Software Defect Prediction for High-Dimensional and Class-Imbalanced Data," 23rd International Conference on Software Engineering & Knowledge Engineering, Eden Roc Renaissance, Miami Beach, USA, 2012.

[9]. K. Purswani, P. Dalal, A. Panwar, and K. Dashora, "Software Fault Prediction Using Fuzzy C-Means Clustering and Feed Forward Neural Network,"

International Journal of Digital Application & Contemporary research, vol. 2, issue 1, 2013.

[10]. S. Haykin. *Neural Networks: A Comprehensive Foundation* (2nd Edition). Prentice-Hall International, 2013, pp. 43-45.

[11]. Neural Networks Toobox: User's Guide, The Mathworks, Inc., Natic, MA 01760-2098, 1992-2002.

[12]. Liu, *Radial Basis Function (RBF) Neural Network Control for Mechanical Systems.* Springer, 2013.

[13]. E. Praynlin, and P. Latha, "Performance Analysis of Software Effort Estimation Models Using Radial Basis Function Network," *International Journal of Computer, Information,* Systems and Control Engineering, vol. 8, no. 1, 2014.

International Journal of Computer Applications Technology and Research
Volume 5– Issue 5, 266 - 274, 2016, ISSN:- 2319–8656

Adoption of Web 2.0 Tools as Learning Instrument in Tanzania Higher Education. Where are WE?

Ashiraf H. Abeid
Mzumbe University
Mzumbe - Morogoro
Tanzania

Abstract: This research paper investigated the awareness, the use and factors hindering the adoption of web 2.0 tools for academic knowledge creation and sharing in higher learning institutions in Tanzania. Structured questionnaires were used to collect data from the targeted respondents (i.e. lecturers and students). Findings from the study revealed that majority of both students and lecturers are very much aware of the different web 2.0 tools, however the academic use of web 2.0 tools is still very limited as opposed to developed nations. This implies that adoption and use of web 2.0 tools still have a long journey before holding its ground in higher learning institutions in Tanzania. Factors hindering the adoption and educational use of web 2.0 tools were established in the study whereby the critical issue identified was the lack of institutional support and strategies to use web 2.0 tools among the stakeholders. It was therefore recommended that institutions formulate policies and strategies that will guide and support the adoption and uses of web 2.0 tools in the higher learning environment because both students and lecturers were very much interested to learn and adopt web 2.0 tools for education use.

Keywords: Adoption, Web 2.0 tools, higher education.

1. INTRODUCTION

Since the introduction of internet technology there has been many efforts to establish various internet application tools for different purposes. A good example of these applications is Web 2.0 tools. Different researchers and institutions have been so much interested on these applications platforms and they have collected and analyze data and tried to recommend the proper use of these technologies. Higher learning institutions are not in exception; they are increasingly researching on the use of new technology (including Web 2.0) in the teaching and learning process in order to enhance and simplify it. Many studies on the use of Web 2.0 tools in education have been done especially in the developed nations (Europe, America and Australia). The concern of this research was to do a very similar study in developing countries (in particular Tanzania) to see whether it will yield to similar or different results which can be generalized for that purpose. A survey instrument was designed to asses and analyzes the use of web 2.0 tools from the perspectives of the Tanzanian students and lecturers.

2. LITERATURE REVIEW

The inception of Web 2.0 in the year 2004 marked the beginning of major changes in the manner and ways people used to interact and communicate over the World Wide Web. Different authors have tried to explain the meaning of the term web 2.0 technology as follows: Web 2.0 is an umbrella term which comprises a number of internet applications tools such a wikis, blogs, audio/video podcasting, really simple syndication, (RSS), and social networking, [2]. Those tools normally allow a two interactive and collaborations among users on the World Wide Web. Another definition provided by [7], Web 2.0 was explained as "a new way to use existing Internet technologies such as XML and JavaScript to enable participation, interaction and collaboration among users, content providers and businesses, rather than just the traditional viewing of static Web pages". The two definitions by the two different scholars above have got some basic common characteristics The key features which can be picked from the definitions above are: Web 2.0 use internet to provides interactive and collaborative platforms among users, Web 2.0 allows users to manipulate and make changes of the data in the web, Web 2.0 provides a participatory platforms in which users can add or edit contents and lastly Web 2.0 provide a user friendly environment. These features differentiated Web 2.0 from the previous version of the web.

The important features of Web 2.0 tools described earlier continuously provide opportunities for the users to improve the process of knowledge exchange as such it allow the accessing, creation, reviewing, editing and dissemination of user generated content, [14]). In the previous version of the Web (i.e. Web 1.0) these features were absent and hence interactions between users were absolutely impossible and information mainly flew from one direction only. The introduction of Web 2.0 tools brought revolutions, It is is no longer a question of one way communication but a two ways of communications among the stakeholders. [14] noted that the innovative nature of Web 2.0 technologies has drastically transformed the manner in which organizations and individuals access, consume and publish information on the web.

There are many applications of Web 2.0 tools, marketing people use it as a marketing tool to promote and communicate

products and service they offer to their customers. So, due to the innovative nature of Web 2.0 tools, there is a substantial evidence for the adoption of Web 2.0 tools among business organizations, in order to enhance communications of their products and services to their stakeholders. For example, [8], in a global survey of the applications of Web 2.0 technologies, shows that there is a high rate of adoption of Web 2.0 technologies by organizations to enhance and sustain their competitive position. In that study, it was observed that things like marketing and promotions of products and services can now be done using some Web 2.0 tools such as social media. In some instance individuals and organizations use web 2.0 tools for formal/informal communications among each other for sharing some important information related to real life experiences and business information. Much as Web 2.0 tools are applied in various areas they are very useful to support the learning and knowledge sharing process in higher education. Advantages of using Web 2.0 tools for supporting teaching and learning process are very obvious: First most of the tools and features of Web 2.0 tools are easy to use and so they do not require one to be a computer/ information and communication technology, (ICT) expert. Only a moderate knowledge of ICT is sufficient for one to use the different Web 2.0 tools. Other advantages of Web 2.0 tools are highlighted as follows: It can increase student's content production, it increase access to classroom resources at any time, it also increase student's motivation and confidence, it enable and increase students to students support and collaborations, it increase student engagement and stimulate collaborative learning. The most important advantage of Web 2.0 tools is that it encourages two ways interactions among students and lecturers beyond the classroom hours. Adoption and uses of social media and other web 2.0 tools for social interactions is growing among people of different ages [14] but its adoption and use by college students and staffs is still very low [2]). This research investigated why?

Different researchers have done surveys from different parts of the world and came up with different results. For example in Australia a research by [3] on the influence of Web 2.0 tools in tourism industries indicated that more and more companies in the tourism sector use varieties of Web 2.0 tools to market their services such as hotel booking and travelling arrangements. This enabled them to save time, cost and improve their performance and it is more convenient and preferred by their customers.

[4] and [5] researches which investigated the use of Web 2.0 technologies by students to enhance academic collaborations in Australia. In the first case they investigated the use of emerging technologies (including Web 2.0) by students. Contrary to what has been said in technology research literatures (e.g. in [11]) students were found not to be active users of technologies to facilitate knowledge sharing in their studies. Specific to Web 2.0 technology it was observed a low usage of web 2.0 tools to enhance knowledge sharing. Despite the fact that majority of students own modern digital devices such as laptops, Tablets and 3G mobile phones/smart phones yet very few of them use them to use Web 2.0 tools for academic knowledge sharing. It was further revealed that awareness of some of Web 2.0 tools such as podcasts was still very low. In the second study [5] investigated the differences between staff and students use of technology in teaching and learning. The results show that only in some of the technologies e.g. gaming, mobile use and web music categories students were very active where as in other categories of technologies such as Web 2.0 the magnitude of the difference was very low, indicating a moderate use of Web 2.0 tools for knowledge sharing. Nevertheless the level of familiarity of Web 2.0 technology among students and staffs was somewhat there but not very much. In Malaysia, [15] investigated the use of Web 2.0 tools by Malaysian students. Their study came up with very similar results as those found by [4] and [5]. That is low use of Web 2.0 tools by students in higher education for academic knowledge sharing, although, in Malaysia it was found that there was very high awareness about Web 2.0 tools and students possesses higher computer knowledge. [9] did a case study research in Malaysia to determine the impact of multimedia learning technologies and Web 2.0 and the way it can be used to promote active collaboration in learning environment. It was revealed from the study that students were so active and successful in using Web 2.0 tools especially blogs for creating a collaborative learning environment between themselves and their lecturers, furthermore blogs helped them to improve their team working and communications skills. These results contravene those founded by [4] and [15]. However precautions should be taken when comparing those results. This is because the study by [9] was administered to a group

of second year students taking a subject "multimedia technologies" in which students were coached, guided and helped to create blogs and websites as part of the course assessments structure which could be the reason for active utilization of Web 2.0 tools that was observed.

Other studies exploring the use of Web 2.0 technologies in higher education include [2] who investigated the use of Web 2.0 technology and its implications to higher education sector in the west. Findings of his study showed that some Universities in the UK and USA use some Web 2.0 tools such as wikis, blogs and bookmaking to facilitate the exchange of academic information between students and staffs. There were also some cases in UK Universities where students and staffs were assisted in using some of these tools for knowledge sharing purposes. Students and staffs were very active and use Web 2.0 tools to stimulate the learning process among themselves. Another study was done by [6] investigated students' perceptions on the usefulness of Web 2.0 technologies in learning. This study was conducted to undergraduate students from different disciplines of studies. Despite the fact that the response of most students indicated high awareness on Web 2.0 tools yet their experience of using Web 2.0 tools to enhance knowledge sharing is still lacking among them as most of them confirmed that they have never used web 2.0 tools for academic knowledge sharing. These results also contradict with the results obtained by [2]. Another study by [1], examined the awareness of faculty members on the benefits of using Web 2.0 tools to supplement traditional teaching methods in the United States. The findings revealed that majority of participants indicated very high awareness and appreciated that Web 2.0 tools are useful to provide an active and collaborative learning environment. These results are in line with those found by [6].

[12] and [13] examined the use of social computing to foster lifelong learning and the impact of social media on learning respectively. In the first case their study showed that Universities in UK have took initiatives to implement social computing applications such as wikis, blogs, social networking systems to facilitate knowledge sharing and collaborations among students, lecturers and other administrative staffs. Some cases of lack of interests among users were also discovered as such some web accounts owned by students were found to be inactive. In the second study there were almost similar observations that some social computing technologies have been used but not so widely applied in education. Thus the common findings on the two studies were *high awareness on Web 2.0 tools* but *limited* ***interests*** on using them for educational purpose. Looking at many technology management literatures awareness and usage of new technologies especially among youths is seen and reported to be high, [11], but the real situation seems to be a different stories. See for example [12] research results which indicate that the number of youths adopted and use social media had increased drastically by the year 2007 and 2008 but when you compare their use for education purposes is still very low. What does this mean then? This means that more specific studies on the use of Web 2.0 in the education sector will help to generate a wider understanding of the real situation on the use of Web 2.0 technologies that could help to improve the current situation as most studies reviewed indicated that the level of Web 2.0 tools applications for academic knowledge sharing is still low.

2.1 The Research Problem

Many previous researches investigated the use of web 2.0 technology found that the level of awareness of web 2.0 technology is high, [15] and [9] However the high level of awareness of web 2.0 tools does not match with the extent of their use especially for creation and sharing of academic knowledge, [9]. This indicates that Web 2.0 tools are not utilized to their full potentials to enhance knowledge sharing among students and lecturers in higher education. Furthermore studies to investigate why the use of Web 2.0 tools to enhance knowledge sharing in higher learning institutions are still limited, many studies were done in Europe, America, Australia and Asia while very few such studies has been conducted for developing countries context especially Africa. This study is an attempt to fill this gape.

2.2 The Research Objectives

The main objective of this research paper is to investigate the adoption and utilization of Web 2.0 for the learning process by students and faculty members in higher learning institutions in Tanzania.

The specific objectives are:

a) To examine Web 2.0 tools awareness and skills among students and lecturers.

b) To analyze the extent of adoption and use of Web 2.0 tools for academic purposes.

c) To analyze factors hindering the use of web 2.0 tools by students and lecturers.

3. RESEARCH METHODOLOGY

To achieve the objective of the study the following method was used. First of all literature review was done to assess the adoption and use of Web 2.0 technologies especially to facilitate teaching and learning and the overall process of academic knowledge sharing. This provided insights about Web 2.0 tools adoption and use in higher education from different areas. Then a survey was adopted in order to gather empirical evidence about the use of Web 2.0 technologies by students and lecturers in higher learning institutions. The design of the survey instrument (i.e. questionnaires) was developed based on earlier administered instruments in order to ensure the validity of the survey. Together with that also a pilot study was done to few selected students and lecturers and after satisfactory review then the questionnaires were administered to the targeted respondents. The University staffs surveyed in this study are full time employed academic staffs who have been involved in teaching responsibilities in those institutions. Students who participated in the study were distributed from different levels of studies from undergraduate to postgraduate levels and they were also chosen from different disciplines of studies. Three higher learning institutions were involved, two public universities and one private university. A total of 800 questionnaires (i.e. 200 for lecturers and 600 for students) were distributed among the respondents from different profession. Stratified random sampling technique was adopted whereby the population was first categorized into different subgroups and then randomly selected from each subgroup until a targeted number is reached. 107 and 496 valid responses from lecturers and students respectively were collected.

4. FINDINGS AND DISCUSION.

4.1 Awareness of Web 2.0 Tools

The responses collected from lecturers and students to assess the awareness of different Web 2.0 tools are shown in table 2 . It can be noted that majority of lecturers and students are aware of the different web 2.0 tools, except that both are less aware of RSS Feed and Mashups. In addition to that students are also less aware of bookmarking. Overal it is convincing that a large number of both students and lecturers are aware of Web 2.0 tools.

Table 1: Awareness of different Web 2.0 tools:

Web 2.0 Tools	Lecturers		Students	
	YES	NO	YES	NO
Blogs	93 (86.9%)	14 (13.1%)	454 (91.5%)	42 (8.5%)
Flickr	56 (52.3%)	51 (47.7%)	97 (19.6%)	399 (80.4%)
Bookmaking	61 (57%)	46 (43%)	245 (49.4%)	251 (50.6%
SNS	101 (94.4%)	13 (12.1%)	402 (81%)	94 (19%)
Wikis	105 (98.1%)	2 (1.9%)	409 (82.5%)	87 (17.5%)
Mashups	48 (44.9%)	59 (55.1%)	210 (42.3%)	286 (57.7%)
RSS Feed	11 (10.3%)	96 (89.7%)	188 (37.9%)	308 (62.1%)
Podcasting	87 (81.3%)	20 (18.7%)	382 (77%)	114 (23%)
Tagging	94 (87.9%)	13 (12.1%)	428 (86.3%)	68 (13.7%)

4.2 Skills on How to Use Web 2.0 Tools

Respondents were then asked about their skills on how to use different Web 2.0 tools. Social Networking Sites, SNS, blogs, wikis, podcasting and bookmarking were the leading known web 2.0 tools among lecturers with a mean score values of above 4. (see table 2). These were followed by flickr and tagging which scored an avaerage 3+. RSS feed and mashups have the lowest mean score of 2+. This indicated that lecturers have high skills in most of web 2.0 tools and minimum skills on how to use RSS feeds and mashups.

Table 2: Lecturers skills on how to use different Web 2.0 tools

Tool	Excellent	Above average	Average	Below average	Very poor	Mean
Blogs	45 (42%)	34 (31%)	22 (21%)	6 (5%)	0 (0%)	4.1
Flickr	32(30)	40 (37%)	11 (10%)	6 (6%)	18 (17%)	3.6
Bookmaking	59 (52%)	23 (22%)	16 (15%)	0 (0%)	9 (8%)	4.1
SNS	69 (64%)	22 (21%)	13 (12%)	2 (1%)	1 (0.9%)	4.5
Wikis	59 (55%)	19 (18%)	21 (19%)	6 (5%)	2 (1.9%)	4.2
Mashups	21 (19%)	16 (15%)	11 (10%)	39 (36%)	20 (18%)	2.8
RSS Feed	39 (3%)	12 (11%)	18 (16%)	27 (25%)	11 (10%)	3.4
Podcasting	57 (53%)	20 (18%)	16 (15%)	9 (8%)	5 (5%)	4.1
Tagging	46 (43%)	24 (22%)	12 (11%)	18 (17%)	7 (6%)	3.7

Table 3: Students skills on how to use different Web 2.0 tools

Tool	Excellent	Above average	Average	Below average.	Very poor	Mean
Blogs	211 (45%)	106 (21%)	89 (17%)	49 (9%)	41 (8%)	3.5
Flickr	53 (11%)	65 (13%)	76 (15%)	168 (34%)	134 (27%)	2.5
Bookmaking	121 (24%)	72 (14%)	45 (9%)	122 (25%)	136 (27%)	2.8
SNS	286 (58%)	77 (15%)	83 (17%)	27 (5%)	23 (4%)	4.2
Wikis	203 (41%)	98 (20%)	83 (17%)	72 (14%)	40 (8%)	3.7
Mashups	61 (12%)	63 (13%)	49 (10%)	108 (22%)	215 (43%)	2.3
RSS Feed	119 (24%)	79 (16%)	48 (10%)	114 (23%)	136 (27%)	2.8
Podcasting	177 (36%)	80 (16%)	132 (26%)	84 (17%)	23 (5%)	3.6
Tagging	35 (22%)	62 (12%)	29(5.8 %)	33 (7%)	14 (3%)	1.1

From the student's side SNS, blogs, wikis and podcasting were the leading known web 2.0 tools with a mean score above 3. They are followed by bookmarking, tagging and RSS feed with a mean score of about 2. Mashups is the last one with a mean score of 1.1. From these findings we can simply conclude that majority of lecturers and students at least have some skills on how to use different web 2.0 tools. Only few tools RSS feed and mashups (for lecturers) and mashups and bookmarking (for students) indicating minimal skills among respondents. Although respondents (both students and lecturers) indicate such minimal skills on those few tools but it is important to reiterate that most of Web 2.0 tools are very simple to learn and use, it is only a question of being ready and support.

4.2 Uses of Web 2.0 Tools

Web 2.0 tools can be used for different purposes, in the education setting as well as other non academic purposes. Respondents of this study were asked to indicate the different ways where they use web 2.0 tools and the results were as follows:

About the use of web 2.0 tools to *search and download academic materials* majority of both lecturers (96%) and students (80%) agreed that they always or often use Web 2.0 tools for that purpose, none of the lecturers and few students rarely or never use web 2.0 tools for that purpose.

Table 4: Lecturers' uses of different Web 2.0 tools

Possible use of web 2.0 tools	Always	Often	Sometimes	Rarely	Never
Search and download academic materials	90 (84%)	12 (11%)	5 (5%)	0 (0%)	0 (0%)
Facilitate online academic discussions	14 (13%)	21 (20%)	17 (16%)	18 (17%)	37 (35%)
Research	97 (91%)	10 (9%)	0 (0%)	0 (0%)	0 (0%)
Update on related topic of interest	84 (79%)	9 (8%)	14 (13%)	0 (0%)	0 (0%)
Online submissions of assignment/papers	54 (51%)	18 (17%)	9 (8%)	16 (15%)	10 (9%)
Contents creations and sharing with others	34 (32%)	16 (15%)	22 (21%)	12 (11%)	23 (22%)
Official/Professional communication	48 (45%)	16 (15%)	21 (19%)	22 (21%)	1 (1%)
Social communication	82 (76%)	11 (10%)	8 (8%)	4 (4%)	2 (2%)

Table 5: Student's uses of different Web 2.0 tools.

Possible use of web 2.0 tools	Always	Often	Sometimes	Rarely	Never
Search and download academic materials	298 (60%)	103 (21%)	55 (11%)	31 (6%)	9 (2%)
Facilitate online academic discussions	193 (39%)	106 (22%)	39 (8%)	77 (15%)	81 (16%)
Research	254 (51%)	142 (29%)	79 (16%)	13 (3%)	08 (2%)
Update on related topic of interest	135 (27%)	88 (18%)	92 (19%)	123 (25%)	58 (12%)
Online submissions of assignment/papers	42 (9%)	30 (6%)	29 (6%)	107 (22%)	288 (58%)
Contents creations and sharing with others	81 (16%)	42 (9%)	99 (20%)	135 (27%)	139 (28%)
Official/Professional communication	88 (18%)	100 (20%)	79 (16%)	98 (20%)	131 (20%)

Social communication	348 (70%)	16 (3%)	56 (11%)	30 (6%)	46 (9%)

About the use of *web 2.0 tools to facilitate online academic discussion;* it appears that lecturers and students have conflicting opinions. While about 33% of lecturers always or often use web 2.0 tools to facilitate online academic discussion, about 61% of students agreed that they always or often use web 2.0 tools for that purposes. Majority of lecturers (53%) and minority of students (16%) never or rarely used web 2.0 tools for online academic discussion. The reason for such conflicting opinions could be that few lecturers have already adopted and engaging their students on web 2.0 tools to facilitate academic knowledge sharing while others are less interested about it. Regarding the *use of web 2.0 tools for research purposes;* both lecturers and students had the same opinions, where (99%) of lecturers and (80%) of students agree that they always and often use web 2.0 tools for research purposes only few of them have denied it.

About the *use of web 2.0 tools to update on related academic topics of interest,* here also the opinions from the students and lecturers were the same though only a simple majority of students (45%) and majority of lecturers (87%) agreed that they always and often used web 2.0 tools to update themselves on various academic topics of interests while 38 % of students and none of lecturers said that they rarely or never used it for that purpose.

About the *use of web 2.0 tools for online submission of assignments/papers* here there were different opinions. While 68% of lectures agreed that they use web 2.0 tools for online assignment/papers submission only very few students about 15% use web 2.0 tools for that purpose. Majority of the students (80%) have never or rarely used web 2.0 tools for online submission of assignments/papers. The reason for such difference could be that so far there are only very few lecturers who engaged their students in web 2.0 tools use.

About the use of web 2.0 tools for *academic knowledge creation and sharing* a simple majority of the lectures (45%) and minority of students (25%) use it for that purpose while minority of the lecturers (33%) and majority of the students (55%) have rarely or never used web 2.0 tools for academic contents creation and sharing. Regarding the use of *web 2.0 tools for making official academic communications* it appears that majority of the lecturer (60%) and few students (38%) use it for official academic communications while (22%) of lecturers and (40%) of students rarely or never use web 2.0 tools for that purpose. Lastly is the use of web 2.0 tools for *social communications (non academic uses*) majority of lecturers (86%) and students (73%) do agree that they used web 2.0 tools for non academic uses while minority of lectures (6%) and students (15%) rarely or never used it for that purpose. From the analysed data above the conclusion which can be drawn for both students and lecturers is that the education use of web 2.0 tools is somehow there but still limited. Although most online activities by students and lecturers to some extent involve learning but they are done in an informal way.

4.4 Level of Adoption

Regarding the adoption level of web 2.0 tools for educational purpose both lecturers (66.6%) and students (61.1%) confirmed that web 2.0 tools are still at the early stage of adoption. This was also confirmed from findings from previous studies observed in literature review, thus the use of web 2.0 tools will have to overcome a number of factors in order to take its way as it is in higher education sector of the developed nations.

Table 6: Adoption of web 2.0 tools: (Lecturers and students)

Level	**Lecturers**	**Students**
Early stage of adoption	71(66.3%)	303(61.1%)
Marginal stage level of adoption	22(20.6%)	102(20.6%)
Advanced stage of adoption	5(4.6%)	10(2%)
No comments	9(8.4%)	81(16.4%)

However practical experience shows that the adoption and use of web 2.0 tools to facilitate academic learning and knowledge creation and sharing for both lecturers and students is gaining momentum. The only problem is that there is still an informal adoption and use of web 2.0 tools for academic activities. The use only depends on the efforts of the individual lecturers/students. The use of web 2.0 tools for social (non academic purpose) is observed to be high nevertheless its use for academic purposes is still encouraging. That is to say there is lack of institutional policies, strategies and support on the use of web 2.0 tools to facilitate academic activities.

4.5 Barriers of using web 2.0 tools

The importance of using web 2.0 tools to facilitate and enhance teaching and learning process are very obvious, that is why they have attracted the attention of different stakeholders arround the world. But in most developing countries particularly Tanzania the level of adoption is still very low as discovered in this research. The interest of this research was also to analyse the barriers of of using web 2.0 tools and several areas were asesed. That is internet connectivity, the ownership of electronics devices, other factors such as skills,organizational support, social gape, confidence and copy rights and legal issues**.**

4.5.1 Internet Connectivity

Table 7: Internet connectivity

	Lecturers		**Students**	
	YES	**NO**	**YES**	**NO**
In Campus	99 (92.3%)	8 (7.5%)	436(87.9%)	60 (12.1%)

Off Campus	21 (19.6%)	86 (80.4%)	124 (25%)	372 (75%)

With regard to internet connectivity this was observed not to be a big problem to all categories of respondents as 92.3% of lecturers and 87.9% of students agreed that they have internet connections within the campus. However the problem to both students and lecturers was the absence of internet connectivity when they are outside the university campus. Although this may not sound to be a very big problem but it is important to give it the weight as it deserves given the fact that due to the increase in the number of students enrolled in higher learning institutions in the recent years it is no longer possible to accommodate all the students in campus and thus internet connection outside the campus become crucial.

4.5.2 Ownership of electronics devices

Table 9: Ownership of electronic devices

Device	Lecturers	Students
Personal laptops	100(93.5%)	388(78.2%)
Desktop computers	56(52.3%)	36(7.3%)
Tablets (Ipad, samsung tab etc)	34((31.8%)	71(14.3%)
Digitl Cameras	69(64.5%)	92(18.5%)
Smart phones/phones with 3G/4G and video capabilities	88(82.2%)	401(80.8%)

With regard to the issue of ownership of electronic devices analysis of the results shows that majority of lecturers (93.5%) and students (78.2%) own personal laptops, more than 80% of both lecturers and students own smart phones/phones with 3G/4G and audio/video capabilities. Also a good number of lecturers own digital cameras and desktop computers while for students only a small number own digital camera. It was also found that very few students and lecturers own tablets. The ownership of electronic devices can be a barrier to use web 2.0 tools, but from the statistics obtained from this study (see table 7) above it appears that majority of the respondents own the basic electronic devices which can enable them to access the internet and use various web 2.0 tools, however it should not be completely ignored that there are some few students (the minority) who doesn't own the basic electronic devices which will enable them to be part of this process. So to a small extent ownership of electronic devices is a barrier to use web 2.0 tools because no one should be left behind if we formalize the use of Web 2.0 tools in the teaching and learning process. The institutions are supposed to provide support to this category of respondents especially students.

4.5 Other barriers of using web 2.0 tools

Other barriers of using web 2.0 tools were assessed by using a number of self reflection statements. Respondents were required to indicate if they agree with the stated barriers and the results were as follows:

Regarding the *skills on how to use web 2.0 tools and general ICT skills* it was founded that both students and lecturers have a good command of web 2.0 tools and ICT skills (see table 9/10). Therefore it can be concluded that web 2.0 tools and ICT skills was not a big obstacle as far as web 2.0 tools adoption and application is concerned.

With regard to *institutional support* both respondents strongly agree that there was a lack of institutional support especially the technical and administrative support to the use of web 2.0 tools.

Table 9: Barriers of using web 2.0 tools (LECTURERS)

Barriers of web 2.0 tools adoption.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Insufficient skills (IT+Web 2.0 tools)	1 (1%)	0 (0%)	3 (2%)	22 (21%)	81 (75%)
Lack of Institutional support	76 (71%)	19 (18%)	8 (7%)	4 (3%)	0 (0%)
Lack of confidence	65 (61%)	34 (32%)	5 (5%)	3 (3%)	0 (0%)
Social gap between students and lecturers	76 (71%)	18 (17%)	10 (9%)	0 (0%)	3 (3%)
Copyright and legal issues	24 (22%)	28 (26%)	50 (47%)	5 (5%)	0 (0%)

Regarding the *lack of confidence* there were conflicting opinions where by lecturers didn't agree that lack of confidence is a barrier whereas majority of students consider it to be a barrier so to some extent this was a barrier in using web 2.0 tools. Respondents were then asked about the *social gap between students and lecturers* and the results indicate that the majority of both lecturers and students strongly agree that there is a social gap between staff and lecturers and hence this was a barrier of adopting and using web 2.0 tools (see table 9/10). Some respondents especially students goes further into explaining this by saying "when it comes to communication or interactions among ourselves, we are okay i.e. we are very free to use any form of communication including different web 2.0 tools however when it comes to communication with lecturers you have to be careful because some lecturers are very selective." Some lectures are not flexible to use varieties of technologies including web 2.0 tools for academic interactions purposes.

The other barrier assessed was *copyrights and legal issues*. A simple majority of lecturers and the majority of students agree that it is a barrier. Though respondents identify this as a barrier however there are a big number of electronic academic

materials which are allowed to be used and they can be shared freely for educational purpose thus it may be inappropriate to consider this to be a big obstacle.

Table 10: Barriers of using web 2.0 tools (STUDENTS)

Barriers of web 2.0 tools adoption.	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly Disagree**
Insufficient skills (IT+Web 2.0 tools)	199 (40%)	58 (12%)	63 (13%)	44 (9%)	132 (27%)
Lack of Institutional support	276 (56%)	108 (22%)	83 (17%)	11 (2%)	18 (3%)
Lack of confidence	59 (11%)	24 (5%)	0 (0%)	342 (69%)	71 (14%)
Social gap between students and lecturers	300 (60%)	97 (20%)	22 (4%)	42 (8%)	35 (7%)
Copyright and legal issues	33 (7%)	18 (4%)	70 (14%)	211 (43%)	164 (33%)

It is clearly understood that the introduction of any new thing like technology, to any institution will have to overcome some opponents and it is therefore the responsibility of organizational leaders to manage the changes appropriately in order to overcome this problem. It should clearly be insisted the need to innovate the teaching and learning process in order to make it much more attractive and participatory especially during this time which is said to be the digital era.

5. SUMMARY AND RECOMMENDATIONS:

The results presented in the previous section above highlighted a number of important issues related to adoption and use of web 2.0 tools. The questions that were asked in the questionnaires sought to understand the awareness and skills on various web 2.0 tools analyze the extent of adoption of web 2.0 tools and analyze the factors hindering the adoption of web 2.0 tools among students and lecturers.

In general students and lecturers skills and awareness on web 2.0 tools is very positive as observed that majority of respondents have high skills and are aware of the different web 2.0 tools. However some respondents indicated unfamiliarity with some web 2.0 tools such as mashups and podcasting. This may be due to lack of interest in the use of those particular tools for learning purposes.

About adoption and the use of web 2.0 tools it was established that adoption level of web 2.0 tools is still at the early and informal stage however majority of students and lecturers in one way or the other have used web 2.0 tools for learning and other social purposes. Some findings from previous researches especially in the developed nation's shows that students and lecturers have already enter into the marginal and advanced level of adoption. By making a direct comparison with the results of this study it is very clear that Tanzanian students and lecturers are lagging behind. However personal experiences during data collection process shows that both students and lecturers are very much interested to adopt and use web 2.0 tools for learning purposes. It is therefore upon the institutions especially those which are interested to increase the use of technology in the teaching/learning process. They should focus their attention, efforts and investments to support this area. Particular attentions should be paid to insist the advantages of using web 2.0 tools and other technologies in the general learning process. In case skills and technical support is needed it should be provided.

Regarding the barriers of using web 2.0 tools it is important to note that most respondents i.e. lecturers and students acknowledged the benefits of web 2.0 tools application in education but there are some barriers identified which inhibit them from actively using web 2.0 tools. The major ones were lack of institutional support, the social gap between students and lecturers, resistance to change and lack of innovation. The minor ones were lack of confidence copyrights/legal issues and insufficient skills. Looking at the barriers above one can definitely say that lack of institutional support and strategies to formalize web 2.0 tools application in the teaching and learning process is the major issue of concern, therefore it is recommended that efforts should be increased to facilitate the adoption and use of web 2.0 tools for improving and enhancing the teaching and learning processes.

6. CONCLUSIONS

The results of this research have provided more insights about students and lecturers adoption and use of web 2.0 tools to facilitate the teaching and learning process. The general finding is that both students and lecturers at least are aware and have some knowledge about web 2.0 tools though the adoption and use of web 2.0 tools in education is still very low and informal. The general recommendation is that there should be establishment of institutional strategies to support the adoption and use of web 2.0 tools for the teaching and learning process. Both stakeholders i.e. students, lecturers and the institutions should play a shared responsibility in order to ensure teaching and learning process by using web 2.0 tools yield the intended outcomes.

7. REFERENCES:

[1] Ajjan H. & Hartshorne, R., (2008), *Investigating Faculty Decision to adopt web 2.0 technologies: Theory and Empirical tests*, [online], Internet & Higher education journal, Volume 11 (2) pp. 71 – 80 available from http://webpages.csus.edu/sac43949/PDFs/Faculty%20Decisons%20Web%202.0.pdf

[2] Anderson, P. (2007), What is web 2.0? Ideas Technology and Implications for Education, [online], JISC, Technology and Standards Watch, available from www.jisc.ac.uk

[3] Elton, N.M., (2013), Web 2.0 and its Influence in the tourism sector, European Scientific Journal volume 9 Issue 20 available online from http://eujournal.org/files/journals/1/articles/1565/public/1565-4687-1-PB.pdf

[4] Kennedy, G. et al, (2007), *The net generations are not big users of web 2.0 technologies. Preliminary findings*, [online], Proceedings Ascilite Singapore 2007, pp. 517 – 526, available from http://www.ascilite.org.au/conferences/singapore07/procs/kennedy.pdf

[5] Kennedy, G. et al., (2008), *immigrants and Natives: Investigating the differences between staffs and students use of technologies in learning*, [online], Proceedings of Ascillite, Melbourne, 2008, pp. 484 – 492

[6] Kumar, S. (2008) *Undergraduate perceptions of the usefulness of web 2.0 in higher education, survey development*, [online], University of Florida, USA, Proceedings of the European conference on e-Learning, pp. 306 – 314, available from http://webtools4teachers.yolasite.com

[7] Lawton, G. (2007) Web 2.0 and Security Challenges, Technology News, Published by IEEE Computer Society

[8] McKinsey & Company, (2007), *How business Organizations are using Web 2.0*: A McKinsey Global Survey, [online], available from www.mckinseyquartely.com

[9] Neo,T.K, Neo, M & Kwok, W.J.J., (2009), *Engaging students in a multimedia corporations learning environments: A Malaysian experience*, [online], Proceedings of the Ascillite, Auckland, 2009, p.p. 674 – 683, available from http://www.ascillite.org.au/conferences/auckland09/procs/neo.pdf

[10] Platt, M., 2009, *Web 2.0 in the Enterprises*, [online], Architecture Journal, Volume 12, issue, 2 pp. 1-5, available from http://msdn.microsoft.com/en-us/library/bb735306.aspx

[11] Prensky, M. (2001). Digital Natives, Digital Immigrants, The Horizon, MCB University Press, Vol. 9(5) available online from http://www.marcprensky.com/writing/Prensk%20-%20Digital%20Natives,%20Digital%20Immigrants%20-%20Part1.pdf

[12] Redecker, C., Ala-Mutka, K & Punie, Y., (2010), *The Impact of social media on learning in Europe*, [online], JRC-EC, Institute for Prospective Technological studies, available from http://ftp.jrc.ec/EURdoc/JRC56958.pdf,

[13] Redecker, C.,Ala-Mutka, K &Punie, Y., (2008), *Learning 2.0: The use of social computing to enhance lifelong learning*, [online] Institute for Prospective Technological studies, available from http://is.jrc.ec.eurpa.eu/pages/EADTUpaperFINAL.pdf.pdf

[14] Romero, C. L., Alarcon-del-Amo, M. and Constantinides, E. (2014). Determinants of Use of Social Media Tools in Retailing Sector. *Journal of Theoretical and Applied Electronic Commerce Research, Volume 9, Issue 1* , 44-55

[15] Zakaria, M. H., Watson, J. & Edwards, S., (2008), *Investigating the use of Web 2.0 technology in Malaysia*, [online], Multicultural Education &Technology Journal, 4(1), pp. 17-29, available from, www.emeraldinsight.com/1750-497X.htm

International Journal of Computer Applications Technology and Research
Volume 5– Issue 5, 275 - 280, 2016, ISSN:- 2319–8656

A Novel Automated Approach for Offline Signature Verification Based on Shape Matrix

Sumbal Iqbal Ahmed
Abasyn University
Peshawar Pakistan

Rashid Jalal Qureshi
Emirates Aviation University
Dubai ,UAE

Imran Khan
Abasyn University
Peshawar, Pakistan

Zuhaib Ahmad
Abasyn University
Peshawar Pakistan

Abdus Salam
Abasyn University
Peshawar Pakistan

Abstract: The handwritten signature has been the most natural and long lasting authentication scheme in which a person draw some pattern of lines or writes his name in a different style. The signature recognition and verification are a behavioural biometric and is very challenging due to the variation that can occur in person's signature because of age, illness, and emotional state of the person. As far as the representation of the signature is concerned a classical technique of thinning or skeleton is mostly used. In this paper, we proposed a new methodology for signature verification that uses structural information and original strokes instead of skeleton or thinned version to analyse the signature and verify. The approach is based on sketching a fixed size grid over the signatures and getting 2-Dimensional unique templates which are then compared and matched to verify a query signature as genuine or forged. To compute the similarity score between two signature's grids, we follow template matching rule and the Signature grid's cell are mapped and matched with respect to position. The proposed framework is fast and highly accurate with reduce false acceptance rate and false

Keywords: Authentication, Biometric identification, Feature extraction, Signature verification, Shape matrix.

1. INTRODUCTION

There are many authentication schemes like fingerprints, voice recognition, iris scanning, retina scan, face recognition etc., however, a handwritten signature has already become an accepted proof of identity of the person in our daily life, especially in financial sectors, where a transaction taken on his or her behalf are being authorized via signatures.

A handwritten signature is the scripted name or legal mark of a person's identity, executed by hand. The signature of a person cannot be stolen, however, its forgery is possible after a good practice. To make a duplicate or false copy of a person's signature the forger tries to produce the signature as closest to the original by carefully learning the basic style of writing and shape characteristics of a signature. Also, the signature recognition and verification are a behavioral biometric and is very challenging due to the variation that can occur in person's signature because of age, illness, and emotional state of the person. So, it is needed to design a system that verifies the signature of a human automatically. It can be operated either as "off-line" signature verification or "on-line" signature verification. In an off-line signature verification, the user write a signature on paper which is then digitize by an optical scanner or a camera, and the biometric system identifies the signature by analyzing its shape. The static information derived in an off-line signature verification system may be global, structural, geometric or statistical. On contrary, in an On-line signature verification, data records the motion of the stylus when the signature is produced, and includes location, and possibly velocity, acceleration and pen pressure, as functions of time. Processing of offline signature is complicated because no stable dynamic features are available and segmenting the signature strokes is very difficult because of variation in writing styles of each person and the variation that can occur in person's signature. Therefore, main challenging problem in design of an offline signature verification system is the phase of extracting features that distinguish between forged and genuine signatures. In this paper, a novel method based on shape matrix is proposed. The proposed technique of signature verification is simple and quick with an excellent recognition rates.

This paper is organized as follow: Section 2 is about the literature review and survey of previously presented techniques for signature verification. Section 3, introduces the proposed model and pre-processing used. Section 4, explain the use of shape matrix in the domain of signature. Section 5, presents the results of our method and its comparisons with the existing similar approaches. Section 6, briefly conclude this research work and point out some possible future work.

2. RELATED WORKS

Many research works on signature verification have been reported. Researchers have applied many technologies, such as neural networks [1] [2], Hidden Markov Model [3] and Support Vector machine [4] and pixel based [5] processing to the problem of signature verification and they are continually introducing new ideas, concepts, and algorithms.

As far as the representation of the signature is concerned a classical technique of thinning or skeletonization is used mostly. It consists of representing the original strokes of the signature by a thinnest representation but preserving the topology or basic structure. After thinning a set of idealized lines is obtained which is called the skeleton or medial axis.
In [6] the authors have used a similar approach of obtaining the skeleton of a signature by using thinning technique. Each pixel that belongs to the signature is then studied and the extraction of endpoints from the signature geometry is done. These endpoints were then joined to draw a closed shaped polygon. The structural features of the polygon like area, perimeter, circularity measure, rectangularity measure, and minimum enclosing rectangle were taken into consideration. A verification function was built by the combination of these features and then its evaluation was done by Euclidean distance matrix.

On a similar note, [7] used signature's skeleton and produced a Delaunay triangulation with selected end points and intersection points. The technique is based on matching the triangles based on relative areas of the triangles and their mutual angles with neighboring triangles.
Kumar et al. [8] presented the technique of Off-line Signature Verification Based on Fusion of Grid and Global Features Using Neural Networks. The skeletonized image is divided into 120 rectangular segments (15x8), and for each segment, the area (the sum of foreground pixels) is calculated. The resulting 96 values form the grid feature vector. Some common global features such as Aspect Ratio, Signature height, Image area, pure width and height were used. The standard back propagation neural network classifier for verification is used. Multilayer feed forward artificial neural network for verification of off-line digitized signatures is used. The proposed NN consists of 30 input variables which are extracted from signature features, and it is designed to verify one signature at a time. Back propagation algorithm is used for training.

All the known thinning techniques generally produce rough branches at the crossing points and junction of lines, which are called bushes. It is fact that two signatures of the same person do have some variations. These small changes in lines can disturb the skeleton a lot and the features based on skeleton points will not be same. This leads to low recognition rate. A technique for Off-line Verification of signatures using a collection of simple shape based geometric features was presents in [9]. The geometric features that are used are a Centre of gravity, Area, Eccentricity, Kurtosis, and Skewness. ANN (artificial neural network) was used to confirm and classify the signatures. In [10] another method for the classification and verification using feature point is presented. The scheme is based on selecting 60 feature points from the geometric center of the signature and compares them with the already trained features points.

In [11] the topological and texture features are extracted from the actual signature set. The system is trained by using these features. The mean feature values of all the actual signature features are calculated. This mean features acts as the model for verification against a test signature. Euclidian distance between template signature features and claimed signature features serves as a measure of similarity between the two. In pre-processing, the binary signature image is cropped to keep only the signature as the content, without noise removal, the features that are based on pixels can results in wrong selections. The system performance deteriorates in case of skilled forgeries too.

Roy et al [12] presented a method of handwritten signature verification using grid based approach and pixel oriented approach. Intersecting points and centroids of two equal half of the signature is being calculated and then those centroids are connected with a straight line and the angles of these intersecting points with respect to the centroids connecting lines are calculated. However, the framework does not produced impressive results, works average in simple forgery to produce a low FAR but skilled forgery case produces 20% FAR.

3. PROPOSED MODEL AND IMPLEMENTATION

The steps involved in the proposed strategy for off-line signature verification based on shape matrix are shown in the block diagram (e.g. Figure 1). The approach is based on sketching a fixed size grid over the signatures and getting a 2-Dimensional simplified templates which are then compared and matched to verify a query signature as genuine or forged.

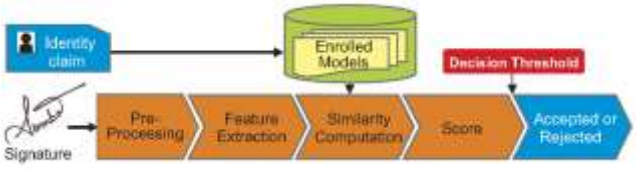
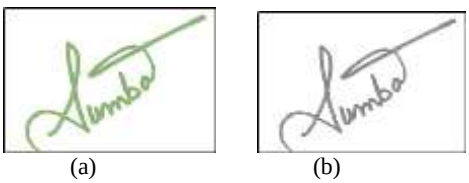


Figure 1 . Proposed Methodologies

3.1 Pre-processing

3.1.1 Converting colour image into binary

In contrast to colour or grey image, handling the binary image is easier and simpler because the image will be in 2-bit representation. Therefore, the colour image is first converted into grey image by setting the intensities of red, green and blue in 30%, 60% and 11% respectively. Then, the Otsu's method [13] was used, which chooses the threshold to minimize the intra-class variance of the black and white pixels. The image obtained after this will be a 2-Dimensional image in which black colour is represented by 0 value (the signature strokes) and white colour is represented by 255 value – the background. However, the inverted image is needed for further processing therefore the image is inverted that is the black pixels are turned into white pixels and vice versa (e.g. Figure 2).



(a) (b)

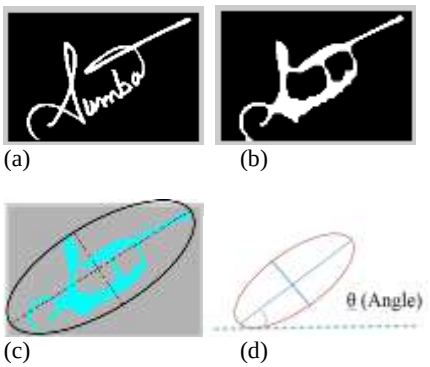
(c) (d)

Figure 2. a) Scanned image b) Gray scale image c) Binary image d) Inverted image

3.1.2 The Slant Removal and Region of Interest (ROI) in the Signature Image:

It is a common and natural that different people sign at different angle and most of the time it is not straight or horizontal to x-axis. But the signature is usually comprised of more than one connected components. We need the signature parts fused together such that we can treat it as one blob or connected area, find its orientation and do the rotation such that it should parallel to horizontal X-axis.

For this purposed, we have used the concept of morphological closing. The orientation angle is found by fitting an ellipse, the major axis of the ellipse shows orientation to X-axis. Once the angle is calculated the image is rotated clock wise or counter clockwise such that it is laying at an angle zero degree or parallel to X-axis as shown in Figure 3.

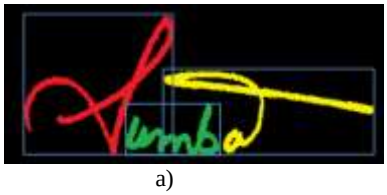


(a) (b)

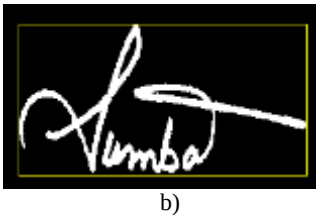
(c) (d)

Figure 3. Slant removal, a) signature b) after morphological closing c) ellipse fitting d) angle computed: 35.1679 °

A signature may have more than one disconnected regions or parts and hence we can have more than one minimum bounding box enclosing each region (e.g. Figure 4a). We need one bounding box that encapsulates the entire signature. To calculate the minimum enclosing rectangle (MER), all the rectangles corners were compared and maximum and minimum x and y coordinates were selected. These values will correspond to top left and bottom right of the minimum bounding box (e.g. Figure 4b).



a)



b)

Figure 4. Region of Interest – ROI, a) MER enclosing each region b) MER enclosing signature

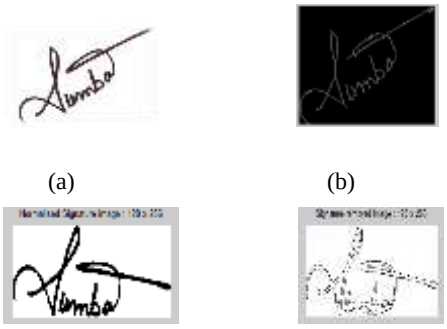
3.1.3 Image Normalization (resizing)

The size of the same person signature may vary. So normalization must be done to scale all the signatures images to a fixed size and minimize the problems that may arise due to difference in size of the signature at the time of comparison. We have used the reference size 128*256 in the proposed approach.

4. SIGNATURE SHAPE MATRIX

4.1 Morphological operators instead of thinning

In our approach, we have not used thinning technique as important information is lost and it produce bushes effect on intersections (e.g. Figure 5b). The great amount of information can be saved by using bridge and remove operator, since the signature boundary is kept saved. Removal of pixels from the boundary of signature can be done but objects are not break apart or separated, (e.g. Figure 5d) shows remove operator's effect.



(a) (b)

(c) (d)

Figure 5. a) Original signature b) Signature image after thinning c) Normalized Signature d) Signature image after applying morphological operation

4.1.2. Signature's Shape Matrix

The proposed signature representation involve division of signature into square grid of size N x N. In this case, we have used N=10. Hence, the signature is divided into a matrix having 10 rows and 10 columns using equal horizontal density method. For each block/cell in the matrix, the signature normalized area is computed. The normalized area is calculated by taking sum of pixels in each box and dividing it by total number of pixels in the signature. The block/cell having area greater than threshold are set to one (1) while the other are set to zero (0). Thus we obtained, for each signature, a binary matrix of 10x10 representing a compact signature code, (e.g. Figure. 6) shows two such signatures and their binary matrix codes.

4.1.3. Similarity Score – Signature's Grid Matching

To compute the similarity between two signature's grids, we follow template matching rule and the Signature grid's cell are mapped and matched with respect to position. For this purpose, simultaneously, the two grids i.e., the Test Signature's grid and the kth signature grid's cells are scanned and compared for a possible match. If the Test signature grid cell's data is equal to the kth signature grid cell data the score is incremented as a step function. The Match Function is

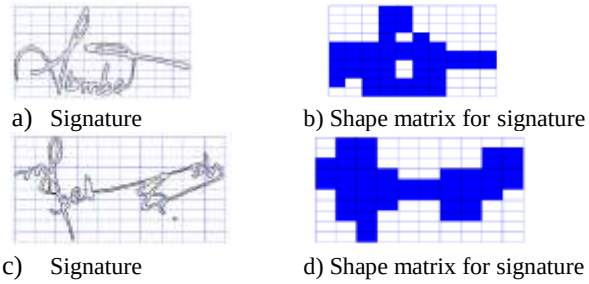
$$Match(i,j) = \begin{cases} 1, & Test(i,j)\ equal\ K(i,j) \\ 0, & otherwise \end{cases} \quad (1)$$

A variable Match (i, j) has a value of one (1) if the data is equal i.e., both grid's cell has a 0 and both grid's cell has a 1. Score is computed as an aggregation function based on the number of times the data in the two grids matches.

$$Score_{t,k} = \sum_{i=1}^{n}\sum_{i=1}^{m} Match(i,j) \quad (2)$$

Where n represents the number of rows and m represents number of columns in the grid. The grid similarity computed in percentage is a normalized score between the two grids showing the percentage match and closeness of the two signature. The Grid similarity is given by

$$Grid_{Similarity\ t,k} = \frac{Score_{t,k}}{n * m} \times 100 \quad (3)$$



a) Signature b) Shape matrix for signature

c) Signature d) Shape matrix for signature

Figure 6. Signature with their binary shape matrix – the template, Grid Matching Score is: 53% Match

We have taken ten (10) sample signatures from each individual and it was encouraging to notice that the scores for the same person were quite higher than the score of similarity between two different person's signatures, it shows the power of discrimination of the proposed approach (e.g. Figure 7).

5. RESULTS AND DISCUSSIONS

Due to unavailability of the standard database, we gathered our own database. This database consists of 1000 signatures that are categorized into 50 classes with each class containing 20 signatures. These are the signatures belonging to 50 subjects and for each person there are 10 genuine signatures and 10 forgery signatures.

Since generally one person can have minor variations in his or her signatures in different trials, each individual was asked to provide 10 samples in multiple sessions over up to 2 weeks period. Ten of the people were requested to give a set of simple forgeries, and the other ten experts were asked to give a set of skilled forgeries to the signature of the 50 subjects.

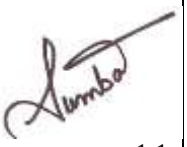
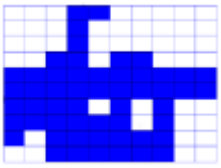
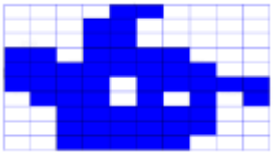
If the grid matching score of the query signature image with respect to models signature image is below the threshold range, the query signature is detected as forged otherwise it is detected as genuine one. There are three different percentages that have been used to measure the performance. These are False Acceptance Rate (FAR), False Rejection Rate (FRR), and Accuracy. FAR is the percentage of forgeries that are incorrectly classified.

$$FAR = \frac{number\ of\ forgeries\ accepted}{number\ of\ forgeries\ tested} \times 100 \quad (4)$$

FRR is the percentage of original signatures that are incorrectly classified.

$$FAR = \frac{number\ of\ original\ rejected}{number\ of\ original\ tested} \times 100 \quad (5)$$

Accuracy is the percentage of signatures which are exactly classified. The Score of similarity in Table. 1 for same person matching is always above 90% an indication that the proposed method is capable of ignoring slight variation that a person usually have among his/her signature.

Signature Image	Signature Shape Matrix	Score with S1-1.jpg
 s1-1		100%
 s1-3		95%
 s1-6		97%

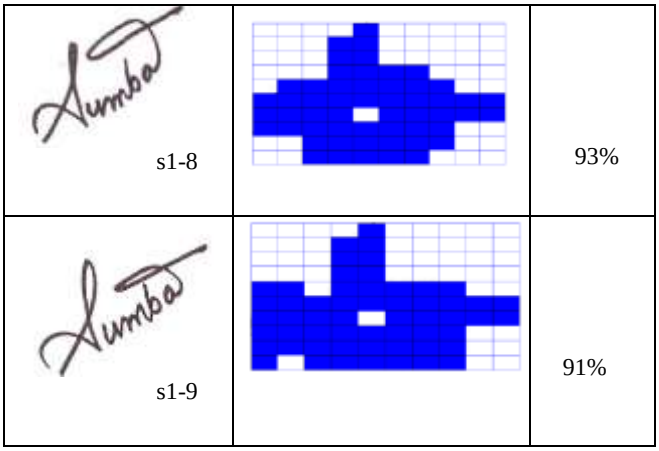


Figure 7. Score of similarity between signatures of same person

Table 1 . Similarity score for the same person's 10 signature

	s1-1	s1-2	s1-3	s1-4	s1-5	s1-6	s1-7	s1-8	s1-9	s1-10
s1-1	100	0	0	0	0	0	0	0	0	0
s1-2	99	100	0	0	0	0	0	0	0	0
s1-3	98	93	100	0	0	0	0	0	0	0
s1-4	95	94	90	100	0	0	0	0	0	0
s1-5	90	91	90	92	100	0	0	0	0	0
s1-6	92	97	94	91	96	100	0	0	0	0
s1-7	92	94	91	92	97	97	100	0	0	0
s1-8	94	97	94	91	96	98	97	100	0	0
s1-9	94	97	96	91	96	96	97	92	100	0
s1-10	94	91	98	91	94	92	95	90	92	100

In contrast, the forgeries scores are in the range of 50% to 60% and can be easily detected by the proposed system (see Table. 2). The Score of similarity for inter-personal matching is always below 50% and in the range of 30 to 40 (Table. 3) shows the power of discrimination of the proposed method.

Table 2. The score of similarity for the same person's 10 forgeries

	f23-1	f23-2	f23-3	f23-4	f23-5
S23-1	55%	50%	57%	51%	56%
	f23-6	f23-7	f23-8	f23-9	f23-10
	58%	53%	51%	54%	51%

Table 3. The score of similarity for the inter-personal variation

	B1-1	B1-2	B1-3	B1-4	B1-5
Test	35%	30%	37%	35%	36%
	B1-6	B1-7	B1-8	B1-9	B1-10
	38%	33%	31%	34%	31%

We have compared our algorithms with two methods one is the 60 feature points Scheme [10] and other is the end point polygon [6].

Our proposed algorithms that were based on structural features got better results than geometric features and end point polygon. The suggested algorithms have successfully rejected skilled forgeries with a very satisfying and good rate and have rejected the simple forgeries very perfectly as shown in Table. 4 and Table. 5. The values of FAR and FRR of our method is better than other schemes.

Table 4. Comparative Analysis of FAR

Forgery Type	60 feature points scheme [10]	End points polygon [6]	Proposed Method (Shape Matrix)
Simple	0.98	0.73	0.65
Skilled	2.08	1.57	1.43

Table 5. Comparative Analysis of FRR

False Rejection Rate (FRR)	
60 feature points scheme [10]	20.83
End points polygon [6]	16.35
Proposed Method (Shape Matrix)	4.15

6. CONCLUSION

Processing of offline signature is complicated because no stable dynamic features are available and segmenting the signature strokes is very difficult because of variation in writing styles of each person and the variation that can occur in person's signature.

The objective of the research was to develop an algorithm which should be relatively fast. We were able to design an offline signature verification system which is highly accurate with reduce false acceptance rate and false rejection rate. The proposed method opens a new area of research by using the original signature stroke and not its thin version or skeleton.

7. REFRENCES

[1] Malekian, V., Aghaei, A., Rezaeian, M. & Alian, M.,2013,Rapid Off-line Signature Verification Based on Signature Envelope and Adaptive Density Partitioning. In IEEE conference on Pattern Recognition and Analysis (PRIA)., pp.1 – 6.

[2] Abdala, M. A., & Yousif, N.M., 2009, Offline Signature Recognition and Verification Based on Artificial Neural Network. In Eng & Tech. Journal.,Vol. 27. No. 7.

[3] Edson, J., Justino, R.,Bortolozzi, F., & Sabourin, R., 2005,A comparison of SVM and HMM classifiers in the offline signature verification. In Pattern Recognition Letters journal. Vol.26 Issue 9. pp.1377-1385.

[4] Ferrer, M. A., Alonso, B., & Travieso, C.M., 2005,Offline Geometric Parameters for Automatic Signature Verification Using Fixed-Point Arithmetic., In IEEE transactions on pattern analysis and machine Intelligence., Vol. 27. No. 6.

[5] Barua, A., Hoque, M.M., Nurul, A.F.M., & Habib, Md.A., 2015, Pixel Based Off-line Signature Verification System. In American Journal of Engineering Research (AJER)., e-ISSN : 2320-0847., p-ISSN : 2320-0936., Vol. 04. Issue-01.pp.187-192.

[6] Zafar, S., Qureshi, R.J., 2009 , Off-line signature verification using structural Features. In FIT '09, 7th International Conference on Frontiers of Information Technology, Abbottabad, Pakistan.

[7] Jan, Z., Muhammad, H., Rafiq,M., & Zada, N., 2015,An automated System for offline signature verification and Identification using Delaunay Triangulation. In Advances in Intelligent Systems and Computing Journal., pp. 653- 663.

[8] Kumar,S., Raja, K.B., Chhotaray, R.K., & Pattanaik, S.,2010,Off-line Signature Verification Based on Fusion of Grid and Global Features Using Neural Network. In International Journal of Engineering Science and Technology.,Vol. 2(12)., pp.7035-7044.

[9] Karouni, A., Daya, B., & Bhalak ,S.,2011,Offline signature recognition using neural networks approach. In Procedia Computer Science world conference on Information technology.,vol. 3.pp. 155–161.

[10] Jena, D., Majhi, Panigrahy,S.K., & Jena, S.K., 2008, Improved offline signature verification Scheme using feature point extraction method. In Cognitive Informatics 7th IEEE International Conference on Cognitive Informatics (ICCI-08). pp.475-480.

[11] Jana, R., Mandal, S., & Chaya, K., 2015, Offline Signature Verification for Authentication. In International Journal of Computer Applications.,Vol. 126 .No. 6. Pp.20-23.

[12] Roy, S., Maheshkar, S., 2014, Offline Signature Verification using Grid based and Centroid based Approach. In International Journal of Computer Applications. Vol. 86. No 8.

[13] N. Otsu. A Threshold Selection Method from Gray-Level Histograms.

International Journal of Computer Applications Technology and Research
Volume 5– Issue 5, 281 - 291, 2016, ISSN:- 2319–8656

Multi agent based Framework for Traffic monitoring in VANET

Divya Chadha
Maharishi Markandeshwar University
Mullana,Ambala
Haryana,India

Reena Dahiya
Maharishi Markandeshwar University
Mullana,Ambala
Haryana,India

Abstract: For last many years studies have been carried out to inculcate technical advancement in the field of VANET traffic monitoring systems. Rapidly growing population and number of vehicles have increased the probability of network congestion, traffic jams and road accidents. To overcome these hazardous situations, series of technologies have been implemented on vehicle infrastructure in recent years. Vehicles are equipped with intelligent on board equipments and communication is facilitated between vehicle and roadside units to enhance road safety. In this paper multi agent base framework is proposed to increase coordination and synchronization among vehicles in VANETs.

Keywords: Mobile Agent, VANET, Vehicle communicator

1. INTRODUCTION

For adding intelligence to vehicles various mechanisms have been proposed so far. Communication between vehicles becomes effective if information sharing and collaboration between nodes is faster. VANET is a subset of MANET. It is necessary for vehicles to communicate with each other and roadside units and share security information and warning alerts. Mobile agents are used in VANETs for attaining this goal. Mobile agents are independent software program that have ability to migrate from one node to another in network during their execution. While migrating and travelling from one node to other mobile agents have capability to collect and deliver information and carry result to multiple locations in network after computations.

Network performance and productivity improves when nodes are equipped with high computational power. Mobile agents can efficiently perform computational task s in dynamic network environment and distributed applications.

Vehicles participating in VANETs have to communicate with one another for sharing various types of information and alerts. For effective communication it is necessary for all the vehicle nodes to participate in communication, respond to neighboring vehicles and follow instructions according to alerts and warnings originated by board units. If nodes are not participating and communicating then it is difficult to achieve coordination and hence enhance network performance. Various researchers have proposed cluster based schemes [1][2][3][4][5][6]. Collision avoidance is necessary for smooth function of system [8][9][10][11][12] have presented collision avoidance based scheme to enhance network performance. Agent based approaches were highlighted in [13][14][15]. Trust establishment and security is crucial feature [16][17][18][19] threw light on these aspects and presented approaches for resolving these issues.[20][21][22]presented overview of routing protocols prevailing in VANETs

Cooperation among nodes is primary feature for attaining effective communication. Nodes may become non cooperative because of various reasons. If driver is not following instructions and alerts, reason may me that he or she is an amateur driver and should not be allowed to drive vehicle. Movements of steering wheel, brakes, speed and acceleration are the important factors that could determine the reasons of instruction violation by vehicles. If alerts and warnings are executed by vehicles they will move in required direction. The violation of rules is determined by a special in proposed work.

Synchronization is also achieved in VANET if desired information is timely shared by all the participating nodes. To achieve this task hierarchy based approach is deployed in proposed framework. Hierarchy table is constructed on the basis of which synchronization and communication priority is established.

In proposed framework it has been assumed that roadside units are fixed after every n kilometer approximately 1 to 2 kilometers. Objective of this approach is to achieve fast synchronization. For achieving this goal cluster file system is used to share copies of tables with all nodes. All the participating nodes are periodically updated with the status of neighboring nodes and hierarchy based approach is used to synchronize nodes and increase cooperation and network productivity. Inside the vehicles various alerts and warning messages are generated to assist driver to take safety actions so that accidents are avoided. These messages are categorized as follows.

- Warning Alerts: Drivers are assisted in driving safely by alerting them through warning signals. In present scenario with latest advancement in automobile designs on board units inside vehicles are designed and programmed to generate alert message. In proposed framework it has been assumed that sensors are embedded on brakes, steering wheel speedometers to assess their performance and check their movements.

Agent software in sensors is programmed to record movements of vehicle and monitors the deviations found in movements. Warning signals are generated in the form of light indication on steering wheel, speedometer, wing mirrors and rear view mirrors and through graphic display on On Board Unit. These signals are of following types:

- Lane Change Warning (LCW): This message is generated to alert driver not to change lane. If direction of vehicle is toward right it intimates and generates alarm to instruct driver for not changing direction to left lane or vice versa. This movement is recorded by agent by determining direction of steering wheel.
- Collision Warning Alert (CWA): Forward collision could be avoided if agents generate warning alert for vehicle moving with high speed with chances of crash or collision.
- Intersection movement Warning (IMW): This warning is originated to alert vehicle to take right path. Intersection points on road are more prone to accidents. If driver is assisted to take right movement on intersections it could add to road safety.
- Turn Assistance Warning (TAW): This alert is generated to assist drivers to take appropriate turn. These warnings suggestions suggest drivers to take appropriate path as per information traffic information received from forward node.
- Safety Alerts (SA): Safety alerts are generated to enhance vehicles as well as its neighboring vehicle's safety. Safety warnings may include road conditions and weather forecast that are generated by sensors embedded in Road side Units. This information is communicated by roadside communicator agents.
- Nearby facility alerts (NFA): Nearby facility information is sometimes urgently required by vehicle and it becomes quite difficult to reach these facilities if their location information is not provided to moving vehicles. Road side units communicate this information to moving vehicles.
- Suggestion alerts (SA): Apart from safety and warning alerts, the proposed framework facilitates vehicles with suggestion alerts so that they can opt right path. This signal is generated through road map displayed on On Board units. Vehicle could move on that path where traffic is less and which is shorter. It shows various path options and benefits as to why that path should be chosen through display units.

2. PROPOSED FRAMEWORK

The proposed framework is divided into two phases. Agent based Vehicle to vehicle (V to V) communication phase and agent based Vehicle to Roadside Unit communication phase. Effective coordination and synchronization will be achieved by communication of these two phases.

2.1 Phase I: Agent based Vehicle to Vehicle communication

For making agent based vehicle to vehicle communication various agents have been inculcated in vehicles. These agents utilize their own information as well as information received from roadside unit. In proposed framework it has been assumed that following agents are embedded at various parts of automobile. This phase consists of two sub phases: Intra vehicle

2.1.1 Intra vehicle communication: vehicle

There are various sensors embedded within vehicle. These are equipped with programmed agents for interaction and recording movements of vehicle. In proposed work it has been assumed that every node will maintain some table with respect to its neighboring nodes. For instance there is a set of n nodes

V_n={V1,V2,V_3,V_4……V_n}. Every i^{th} node will maintain all the tables with respect to its neighbor vehicles. Following tables are maintained in every vehicle's database.

1) NeighborExpected_table
2) NeighborActual_Table
3) Decision_Table
4) Vehicle_Grading_Table

1) NeighborExpected_table: This table is maintained by Vehicular performance agent (VPA) embedded in every node in VANET with respect to all neighbor nodes at specific hop distance. Information regarding neighbor nodes is stored and maintained by every node in this table. In this framework four situations have been described where expected speed and distance values of neighbor nodes are shown in NeighborExpected_table.
2) NeighborActual_Table :This table records actual speed of vehicles by Vehicular performance agent (VPA) and distance with neighbor nodes under various circumstances. Values stored in this table are used to calculate actual deviations. This table is stored in VPA of every neighbor node.
3) Decision_Table :Decision table is described by Vehicular performance Evaluation agent (VPEA) after analyzing the deviations after comparing actual values and expected values. It is important that vehicle should follow the decision for safe and smooth drive. After matching expected values with actual values obtained from vehicle, VPA evaluates the difference and provides decision for smooth driving of vehicle. For instance if it is found that there is danger alarm at lane change then vehicle has to control speed and stop its steering wheel rotation. Such decisions are taken by VPEA to control the movements of vehicle after analyzing its neighbors. It indicates to its own vehicle that neighbor is coming at particular speed at a specific distance and what is expected behavior of vehicle in that situation.

4) Vehicle Grading Table

Vehicle Grading table is maintained by VPEA at every vehicle to grade vehicle performance by finding whether vehicle is following the decision given by Decision_table or not. If vehicle is following indications generated by VPEA then grade is α given to vehicle and if not followed β is grade for that situation.

d by agents inside the vehicle to assess the vehicle performance. Following steps are performed:

- Every ith node will broadcast hello message to all the nearest neighbor nodes.
- Neighbor nodes will reply back to ith node and will send their consent to share information.
- All the nearest neighbors will then share some tables and will maintain copies of all four tables in context of a particular neighbor.

Table 1. NeighborExpected_Table

Neighbor Nodes	Hop Distance	Expected_Distance			Expected _speed		
		Distance_alarm			Density_level		
		Safe>	Alert	Danger	High Density	Average Density	Low Density
Collision	1	10	5-10	3-5	Less than 40	40-60	60-70
Lane change	1-3	20	15-20	10-15	Less than 40	40-60	>= 60
Turn Assistance	1	20	10-20	5-10	Less than 40	40-50	50- 60
Intersection	1	20	10-20	8-10	Less than 30	40-50	50-70

Table 2 NeighborActual_Table

These tables are created and maintaine

	Vehicle_Tags	Hop Distance	Actual_Distance	Actual_speed	Density
V1	V_LEFT1	1	8	60	High
V2	V_Back1	1	30	60	Low
V3	V_Forwad1	1	20	75	Average
V4	V_Right1	1	10	80	Low

Table 4. Vehicle_Grading Table

Neighbor Nodes	Vehicle_Tags	Decision followed V_i					Grade
		Signal	Brake_move _update	Steering_Whl_Rotation_update	Speed_update	Distance_update	
	Turn Assistance						
V1	V_LEFT1	DANGER	No_move	Rotate_Left	Speed_Reduce	Distance_Increased	β
V2	V_Back1	SAFE	No_move	Rotate_Left	Speed _same	Distance_Same	α
V3	V_Forwad1	ALERT	Pushed	Rotate_Left	Speed _same	Distance_Same	β
V4	V_Right1	DANGER	Pushed	Rotate_Left	Speed_normal	Distance_Increased	α
	Collision						
V1	V_LEFT1	DANGER	Pushed	No_rotate	Speed_red	Distance_Increased	α
V2	V_Back1	SAFE	No_move	No_rotate	Speed _same	Distance_Same	α
V3	V_Forwad1	ALERT	Pushed	No_rotate	Speed_Reduce	Distance_Increased	α
V4	V_Right1	DANGER	Pushed	No_rotate	Speed_Reduce	Distance_Increased	α
	Lane change						
V1	V_LEFT1	DANGER	Pushed	Rotate_Left	Speed _same	Distance_Same	β
V2	V_Back1	SAFE	No_move	Rotate_Left	Speed _same	Distance_Same	α
V3	V_Forwad1	ALERT	Pushed	Rotate_Left	Speed_Reduce	Distance_Increased	α
V4	V_Right1	DANGER	Pushed	Rotate_Left	Speed_Reduce	Distance_Increased	α
	Intersection						
V1	V_LEFT1	DANGER	Pushed	Rotate_Left	Speed_Reduce	Distance_Increased	α
V2	V_Back1	SAFE	No_move	Rotate_Left	Speed _same	Distance_Same	β
V3	V_Forwad1	ALERT	Pushed	Rotate_Left	Speed _same	Distance_Same	β
V4	V_Right1	DANGER	Pushed	Rotate_Left	Speed_Reduce	Distance_Increased	α

Neighbor Nodes	Vehicle _Tags	Hop Distance	Distance_ alarm	Speed_ level	Density	Table 3. Decision Table			
						Signal	Brake_move	Steering_Whl_Rotation	Speed
Turn Assistance						Left_Turn/Right_Turn			
V1	V_LEFT1	1	Danger	High	High	DANGER	Pushed	No_Rotate	Reduce
V2	V_Back1	1	Safe	Normal	Low	SAFE	No_move	Rotate_Left/Rotate_Right	No_change
V3	V_Forwad1	1	safe	High	Average	ALERT	Pushed	Rotate_Left/Rotate_Right	Reduce
V4	V_Right	1	Danger	High	Low	DANGER	Pushed	Rotate_Left/Rotate_Right	Reduce
Collision						Left_Turn/Right_Turn/Foward			
V1	V_LEFT1	1	Alert	High	High	DANGER	Pushed	No_Rotate	Reduce
V2	V_Back1	1	Safe	Normal	Low	SAFE	No_move	Rotate_Left/ Right/No_Turn	No_change
V3	V_Forwad1	1	Safe	High	Average	ALERT	Pushed	Rotate_Left/ Right/No_Turn	Reduce
V4	V_Right1	1	Alert	High	Low	DANGER	Pushed	Rotate_Left/ Right/No_Turn	Reduce
Lane change						Left_Lane/Right_Lane			
V1	V_LEFT1	1	Danger	High	High	DANGER	Pushed	No_Rotate	Reduce
V2	V_Back1	1	Safe	Normal	Low	SAFE	No_move	Rotate_Left/Rotate_Right	No_change
V3	V_Forwad1	1	Safe	High	Average	ALERT	Pushed	Rotate_Left/Rotate_Right	Reduce
V4	V_Right1	1	Danger	High	Low	DANGER	Pushed	Rotate_Left/Rotate_Right	Reduce
Intersection								Left_Turn/ Right_Turn/Forward	
V1	V_LEFT1	1	Danger	High	High	DANGER	Pushed	No_Rotate	Reduce
V2	V_Back1	1	Safe	Normal	Low	SAFE	No_move	Rotate_Left/ Right/No_Turn	No_change
V3	V_Forwad1	1	Safe	High	Average	ALERT	Pushed	Rotate_Left/ Right/No_Turn	Reduce
V4	V_Right1	1	Danger	High	Low	DANGER	Pushed	Rotate_Left/ Right/No_Turn	Reduce

Following steps are followed in detail for performing tasks of this framework:

1) Every node maintains NeighborExpected_Table with respect to its neighbor nodes. Given below is neighbor expected table of node V_I for all neighbor nodes surrounding V_i

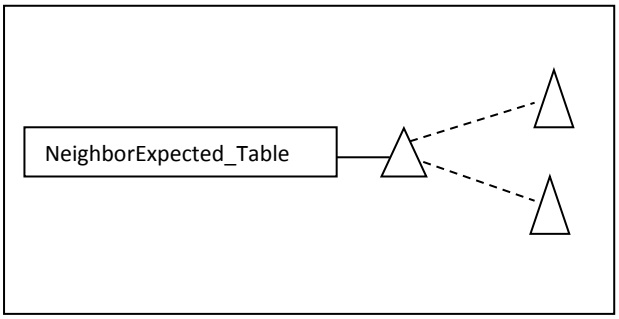


Figure1. NeghborExpected_Table at VCi

2) Node V_i send empty copy of NeighbourActual_Table to all the V_{n-i} neighbor nodes to get actual value of neighbor performance. Neighbor nodes after filling values send NeighborActual_Table to V_i node.

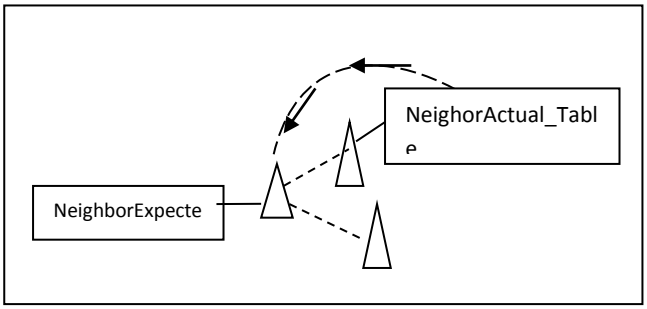


Figure 2. NeighbourActual_Table Sbmission

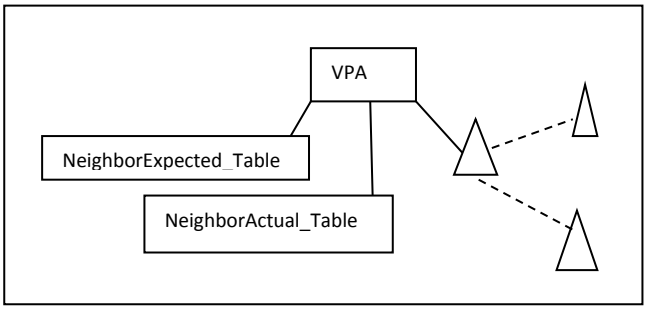


Figure. 3 NeighborActual & Expected Table at VPA

The above figure shows two tables maintained at V_i with respect to all $V_{n-i..}$ All the nodes are equipped with a copy of

NeighborExpected_Table and NeighborActual_Table received from rest of the neighbor nodes.

3) These two tables are registered at Vehicle Performance agent of vehicle and submitted to Vehicular performance evaluation agent to compare the values of two tables and decide what action vehicle has to follow.

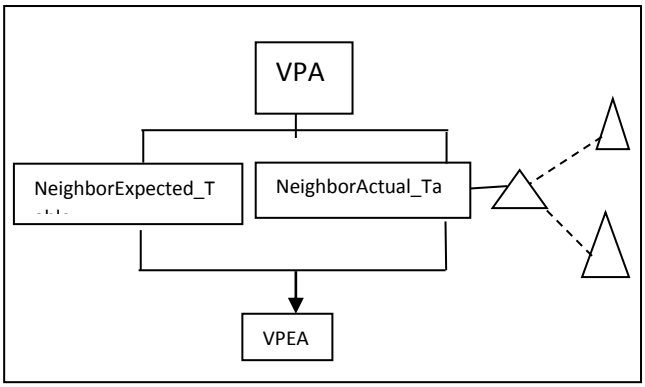


Figure 4. Table submission at VPEA

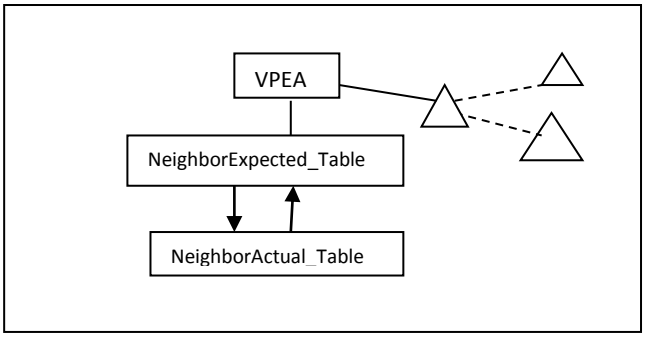


Figure 5. Comparison of Expected and Actual

4) After comparing the values of two tables vehicle performance evaluation agent finds grade of ith vehicle on the basis of fact that whether it has followed the decision or not. It assigns grade in Vehicle Grading table by assessing the performance of vehicle as per decision given by VPEA.

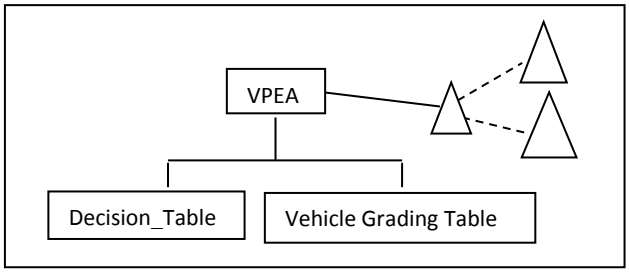


Figure 6. Decision and Vehicle Grading Table

The following algorithms are implemented Vehicle performance evaluation agent (VPEA): This agent is responsible to monitor and assess performance of vehicle by counting the number of times it has obeyed and followed the instructions and alerts. A vehicle is said to be honest and obeying rules if its brake state, steering wheel angle , speed limit changes as required by warning alerts. These alerts are originated in the form of electronic light signals on steering wheel, electronic brakes light and graphic display on On Board Units.

Main responsibilities of VPEA are:

1) To compare the actual movement of parts of vehicle with expected movements and find the deviations.
2) To grade a vehicle on basis of its performance.

Input: NeighborExpected_Table, NeighborActual_Table

Output: Decision_Table_VCi

{ for (i=n-i,,n)

{

do

{

get(density_level(expected_speed),Distance_parameters(expected_distance))←NeighborExpected_Table

get(actual_speed,actual_distance,density← NeighborActual_Table

}

If((Turn_assistance(VC_i))||(Lane_change(VC_i))||(Collosion(VC_i))||(Intersection(VC_i)))

{

If((actual_speed>expected_speed(Density_level))&&(atual_distance<expected_distance(Distance_parameters=Danger))

{

Speed_level=High
Distance_alarm=Danger
{
Set Signal= DANGER
set speed=speed_reduce
Set distance=distance_increase
Set brake_move=pushed
}

}

If((actual_speed>expected_speed(Density_level)|| (atual_distance<expected_distance(Distance_parameters=Alert))

{
Speed_level=High
Distance_alarm=Alert
{

```
Set Signal= ALERT
set   speed=speed_reduce
Set distance=distance_increase
Set brake_move=pushed
}
}

If(((actual_speed>expected_speed(Density_leve)l)||
(atual_distance>=expected_distance(Distance_parameters)))
{
Speed_level=high
Distance_alarm=safe
{
Set Signal= alert
set   speed=speed_reduce
Set distance=distance_same
Set brake_move=pushed
}
If(((actual_speed<expected_speed(Density_level))||
(atual_distance<=expected_distance(Distance_parameters)))
{
Speed_level=normal
Distance_alarm=alert/danger
{
Set Signal= alert/danger
set   speed=speed_reduce
Set distance=distance_same
Set brake_move=pushed
}
}
If((actual_speed<=expected_speed)&&
(atual_distance>=expected_distance(Distance_parameters=safe))
{
Speed_level=Normal
Distance_alarm=safe
{

Set Signal= safe
set   speed=speed_same
Set distance=distance_same
Set brake_move=no_move
}
}

If left_turn(VCn-i)
Set steering_whl_rotation(VCi)=rotate_left
If  right_turn(VCn-i)
Set steering_whl_rotation(VCi)=rotate_right
If no_turn(VCn-i)

Set steering_whl_rotation(VCi)=No_rotate

Create decision_table_VCi

put
(speed_level(VCi),Distance_alarm(VCi),signal(VCi),brake_m
ove(VCi),steering_whl_rotation(VCi))→Vehicle_Grading_Ta
ble

update Decision_table_VCI

}
```

Vehicle performance in terms of vehicle grade could be evaluated as follows:

```
Input: Decision_Table_VCi

Output: Vehicle_Grading_Table

for (i=n-i,n)

{

If(signal=danger)

{

If                    ((brake_move!=brake_move_update)||
(Steering_Whl_Rotation!=Steering_Whl_Rotation_update)||
(actual_speed!=speed_update)
||(actual_distance!=distance_update))

Grade=β

ElseIf          ((brake_move!=brake_move_update)          ||
(Steering_Whl_Rotation!=Steering_Whl_Rotation_update)or
(actual_speed!=speed_update)                              ||
(actual_distance!=distance_update))

Garde=α

}
OR

If(signal=alert)

{

If          ((brake_move!=brake_move_update)          ||
(Steering_Whl_Rotation!=Steering_Whl_Rotation_update)or
(actual_speed!=speed_update)                              ||
(actual_distance!=distance_update))

Grade=β

ElseIf((brake_move!=brake_move_update)                    ||
(Steering_Whl_Rotation!=Steering_Whl_Rotation_update)||
(actual_speed!=speed_update)
||(actual_distance!=distance_update))

Grade=α

}

If (Signal =Safe)

{

If((brake_move!=brake_move_update)||
(Steering_Whl_Rotation!=Steering_Whl_Rotation_update)||(a
ctual_speed!=speed_update)                                ||
(actual_distance!=distance_update))

Grade=β
```

Elseif((brake_move!=brake_move_update)||
(Steering_Whl_Rotation!=Steering_Whl_Rotation_update)||(a
ctual_speed!=speed_update) ||
(actual_distance!=distance_update))

Grade=α}

Update Vehicle_Grading_Table

}

Vehicle communicator agent: This agent communicates status of vehicle performance with neighbor nodes. Nodes near to a particular vehicle follow actions to enhance road safety.

2.1.2 Master vehicle communicator

Master vehicle communicator is cluster head. For first iteration MVC is elected randomly. But for iterations MVC is calculated on the basis of some attributes. MVC election criteria are set by Roadside Unit on the basis of parameters received from MVC. Vehicle communicator of every vehicle calculate grade of vehicle and that information is communicated to other vehicles and MVC records that information and when it leaves the cluster it submit all the current updated information to roadside unit after every n kilometers. The table created by MVC has following attributes:

1) Vehicle grade: Vehicle grade is retrieved from Vehicle_Grading_table and this grade is considered as one of the contributing factor for determining the rank of vehicle. The entire vehicles have to submit this report to MVC. In case vehicle is not submitting the report that vehicle is declared as non participating node and its rank in the hierarchy table get diminished.
2) Participation history: participation history is determined by counting the number of times vehicle has participated in communication and it has followed the instructions. It is counted on the basis of vehicle_grade that is communicated by Vehicle communicator agent. Participation_history_tabl e is maintained at MVC and it record various attributes that are considered in Participation_history_table.
   Participation_History_Table Table 5. is created to determine participation value of vehicle. Participation value is calculated on the basis of Grades achieved by vehicles in different situation. This table is prepared on the basis of vehicle grade from vehicle_grade_table. Alpha is considered as positive contribution and Beta as negative. Participation_History_Table calculates total number of positive and negative contributions of all participating vehicles and enters highest number of contribution in the participation value column.

3) Path selection: Path is selected by a vehicle on the basis of suggestions that it has received from forward vehicle. The vehicle ahead gives path selection suggestion to vehicles following that vehicle. For a particular cluster MVC provides details of safe paths and smooth paths to road side units and RSU agents communicate that information to new MVC that is elected for next cluster and this process goes on.

In figure 7 vehicle nodes are connected through communication channel. MVCA of one cluster C1 is communicating with MVCA of another cluster C2 and sending safe and smooth path information to previous cluster. MVCA of C1 submits information regarding path to nearest RSU while leaving the cluster. RSU uses this message to determine whether nodes are following path or not. Safe and smooth path followed by a particular vehicle is used as one of the parameter for assessing performance of vehicle by RSU.

Table 6 shows grading of path followed by vehicles and this Path_Follow_Grade is helpful in determining whether a vehicle is following a path or not. In case a vehicle is not following path its value will be degraded in final ranking of vehicles.

4) Energy level: Energy level of nodes will also be considered as one of the parameter to calculate hierarchy rank of a particular node.

   MVC of each cluster generates a table that contains overall performance of vehicles in cluster in terms of vehicle grade, participation history, path selection and energy level. This table is submitted to nearest RSU and MVC of preceding clusters and this information is circulated by MVC to all the vehicles of cluster.Table 7 is MVC table given below.

2.2 Phase II: Agent Based Vehicle to roadside communication

Roadside Unit: It has been assumed in this framework that roadside unit is installed after every n kilometers and Road side unit has agent software embedded inside it that is programmed to store information that is submitted by Master Vehicle Communicator agent. MVCA of every cluster submits some information to nearest RSU at the time of leaving a cluster. Roadside Unit agent maintains a table in which it records information regarding all the vehicles in cluster. MVC submits details of MVC_table to the road side unit after every 1 or 2 kilometers.

Table 4.5. Participation_History Table

Neighbor Nodes	Vehicle_Tags	Vehicle Grade					Participation _Value
		Signal	Turn Assistance	Collision	Lane change	Intersection	
V1	V_LEFT1	DANGER	β	α	β	α	2
V2	V_Back1	SAFE	α	α	α	β	3
V3	V_Forwad1	ALERT	β	α	α	β	2
V4	V_Right1	DANGER	α	α	α	α	4

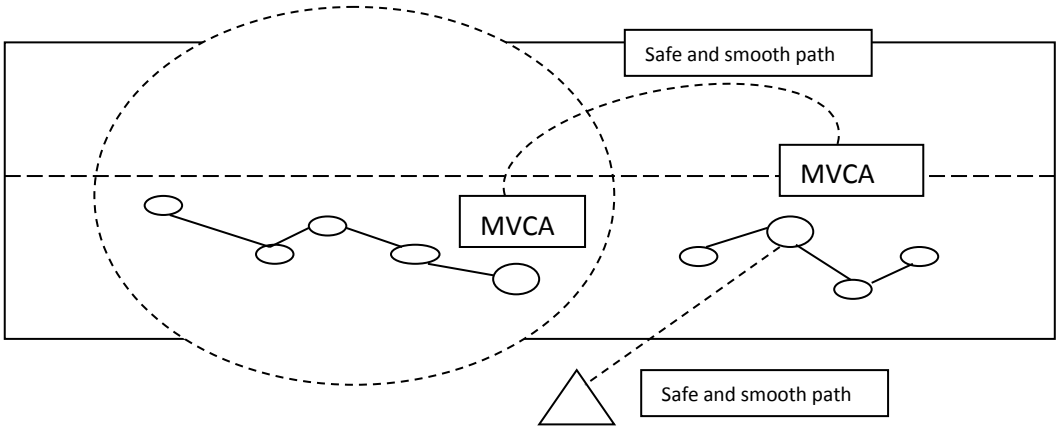


Figure 4.7.Path Selection by MVC

Table 6. Path_Follow Table

Neighbor Nodes	Vehicle_Tags	Smooth_path	Safe_ Path	Actual_Path	Path_Follow_Grade
V1	V_LEFT1	V2_C1 V3_C1 V5_C1 V2_C2 V4_C2	V2_C1 V3_C1 V5_C1 V2_C2 V4_C2	V2_C1 V3_C1 V5_C1 V2_C2 V4_C2	α
V2	V_Back1	V1_C2 V3_C2 V4_C3 V5_C3	V1_C2 V2_C2 V4_C3 V5_C3	V1_C2 V3_C2 V5_C3 V5_C4	β

Table 7. MVC Table

Cluster Nodes	Vehicle_Tags	Vehilce_grade	Participation_History	Path_follow	Energy_level
V1	V_LEFT1	β	2	α	65
V2	V_Back1	α	3	β	65
V3	V_Forwad1	β	2	β	40
V4	V_Right1	α	4	α	60

Hierarchy Determination and MVC Election: On the basis of grade, participation history, path selection and energy level of vehicle hierarchy is determined. There are many benefits of establishing hierarchy ranking of vehicles of VANET. The vehicle having highest rank will have capability to become Master Vehicle communicator or cluster head only. Because that vehicle is efficient in all aspects and will be able to replicate and share copies of updated information more efficiently. This will be beneficial in achieving data synchronization and process synchronization also as all the nodes will be associated very quickly and easily with fast availability of updated information.

Table 8. Hierarchy_Table

Cluster Nodes	Vehilce_ grade	Participation _History	Path_selection	Energy_ level	Hierarchy
V1	β	2	α	65	3
V2	α	3	β	65	2
V3	β	2	β	40	4
V4	α	4	α	60	1

This Hierarchy_table determines the priority of vehicles participating in the cluster. Hierarchy is determined on the basis of vehicle grade, its participation history, path selection and energy level.
This table is maintained at RSU on the basis of MVC. For example in above table vehicle V3 is the vehicle elected as master vehicle communicator for next cluster.

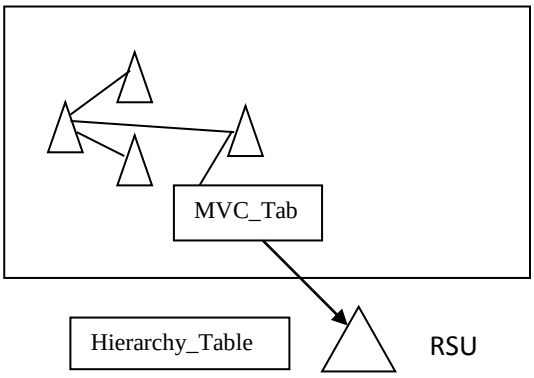


Figure 9. MVC_Table submission and Hierarchy_Table formation at RSU

Master Vehicle communicator submits MVC_Table to nearest RSU. It is not necessary that cluster reformation takes place or not. MVC_Table is submitted to every nearest RSU by Master Vehicle communicator. It is responsibility of RSU to elect the next Master Vehicle communicator on the basis of grades earned by vehicles participating in VANET. RSU determines hierarchy of participating vehicles on the basis of vehicle grade, participation history, path selection and energy level of nodes. Hierarchy table is updated whenever MVC_Table is submitted by MVC at RSU

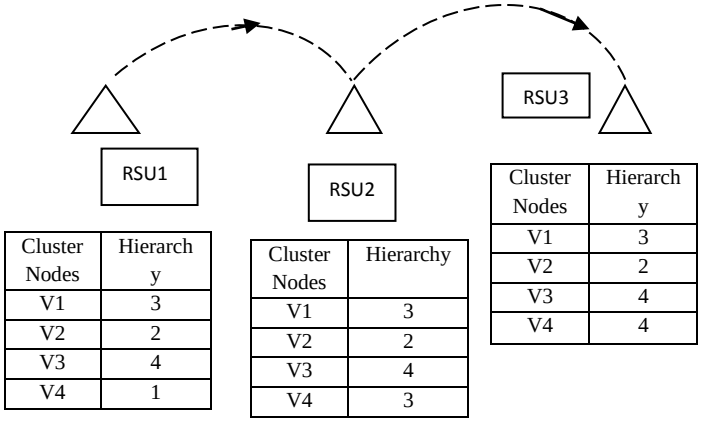
2.3 Phase III: Roadside Unit to Roadside unit communication

RSU maintains copy of all the tables. Vehicle communicators Agent (VCA) of all vehicles submit details of all the four tables to MVC of that cluster. MVC submits all the tables to Roadside Units. Road side Units communicates this information to another RSU so that information about nearby facilities and warnings are communicated to nodes in cluster. All the RSUs receive information from previous RSU and update all the participating nodes with road conditions, Path conditions, nearby facility information, smooth and safe path updates.

2.3.1 Motivation based approach:

Motivation based approach is used in proposed framework to promote vehicles with low grade in hierarchy to participate actively and follow warnings and alerts suggested by VPEA. This process is followed as per below mentioned steps.

1) Roadside units asses performance of vehicles and generates hierarchy. This information is submitted to next nearest RSU e.g. V4 in Hierarchy_Table
2) If next RSU again maintain Hierarchy table and Value of V4 grade increases then V4 will get advantage in terms of discount on any nearby facility.



3) Hierarchy table is transmitted from one RSU to another and when it reaches last RSU3 value of V4 increases.
4) Increasing value of V4 indicates that V4 has started participating in VANET and additional benefits are given to V4. These benefits could be a discount at nearby facility. RSU3 will communicate to nearest restaurant, hospital, filling station or any other nearby facility to provide 10% discount to that vehicle.
5) If a vehicle is not responding and participating in communication and not following any alert or warning,

then this report should be submitted to vehicle recovery agent (VRA) that resides with RSU.

6) VRA plays significant role in detecting problems with non-participating nodes. VRA analyzes Fuel level, vehicle grade that further consist of all aspects that have been considered in all the four tables, Hierarchy value and all the parameters that were actually utilized for its calculation. After analyzing all the factors it decides whether a node is selfish or needy.

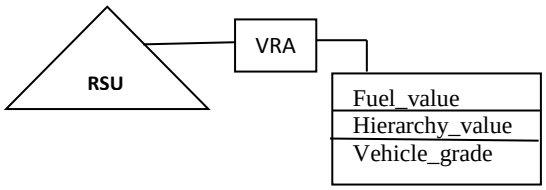


Figure 10. Vehicle_Recovery Agent

7) If a vehicle is rich in all aspects and not participating then the Vehicle_Status is selfish and if it is deficient in any of the three aspect then Vehicle_Status is needy and corrective measures are taken to improve the condition and status of that vehicle. Vehicle status is determined on the basis of Hierarchy_value and Fuel_Value

If {(Hiearchy_value=Low) and (Fuel_Value<=25%)}

Then Vehicle_Status=Needy

Else Vehicel_Status=Selfish.

This Vehicle_Status is reported by VRA to nearby road safety authorities and corrective measures are taken after checking driver details. VRA plays an important role in determining status of vehicle.

3. CONCLUSION

In this work various minor aspects have been presented that could play a significant role in determining the performance of vehicles. Road safety could only be enhanced if vehicles obey rules and run smoothly. In this multii agent based proposed framework vehicle performance is calculated on the basis of various movements. Agent embedded inside vehicles and roadside units are the main entities that enable fast communication among vehicles. Communication among vehicles are only effective if an individual vehicle is able to present its exact and accurate status to surrounding nodes.Agents inside vehicle are calculating vehicle's performance and communicating regularly with other vehicles and road side units. For communication to be effective it is necessary to improve both inter as well as intra vehicle communication. Our future task would be to implement this framework in JADE. After agent communication network performance will be verified in network simulator.

.

4. REFERENCES

[1] Xiaonan W. and Huanyan Q., 2013. "Constructing a VANET based on cluster chains," International Journal of Communication Systems,; doi:10.1002/dac.2484. 16.

[2] Venkataraman H., Delcelier R., and M.Muntean G., 2013. "A moving cluster architecture and an intelligent resource reuse protocol for vehicular networks," Wireless Net-works, Vol. 19, pp. 1881-1900.

[3] Tang X., Sun H., Sun L., Tan C., and Xu G, 2013. "Game theoretical approach for ad dissemination in cluster based VANETs," in Proceedings of IEEE International Conference on Signal Processing, Communication and Computing, pp. 1-6. 18.

[4] Schoch E., Kargl F., Weber M., and .Leinmuller T., 2008, "Communication patterns in VANETs," IEEE Communications Magazine, Vol. 46, pp. 119-125. 19.

[5] Hassanabadi B., Shea C., Zhang L., and Valaee S., 2014. "Clustering in vehicular ad hoc networks using affinity propagation," Ad-Hoc Networks, Vol. 13, , pp. 535-548. 20.

[6] Rawshdeh Y. Z. and Mahmud S. M., 2009. "Toward strongly connected clustering structure in vehicular ad-hoc networks," in Proceedings of the 70th IEEE Vehicular Technology Conference, pp. 1-5. 21.

[7] Girinath D. R. and Selvan S., 2013. "A novel hierarchical model for vehicular traffic regulation," Telecommunication Systems, Vol. 52, pp. 2101-2114

[8] Amudhavel J, et al. 2015.An robust recursive ant colony optimization strategy in VANET for accident avoidance (RACO-VANET). International Conference on Circuit, Power and Computing Technologies (ICCPCT); Nagercoil.. p. 1–6.

[9] Valdes-Vela M, Toledo-Moreo R, Terroso-Saenz F, Zamora-Izquierdo MA2013. An Application of a fuzzy classifier extracted from data for collision avoidance support in road vehicles. Journal of Engineering Applications of Artificial Intelligence.26(1):173–83.

[10] Milanes V, Perez J, Godoy J, Onieva E. 2012.A fuzzy aid rear-end collision warning/avoidance system. Journal of Expert Systems with Applications.; 39(10):9097–107.

[11] Amudhavel J, et al. . 2015.A krill herd optimization based fault tolerance strategy in MANETs for dynamic mobility. International Conference on Circuit, Power and Computing Technologies (ICCPCT); Nagercoil. p. 1–7.

[12] Wang L, Schmidt B, Nee AYC. 2013. Vision-guided active collision avoidance for human-robot collaborations. Journal of Manufacturing Letters.; 1(1):5–8

[13] Nadeem, S. Dashtinezhad, C. Liao, and L. Iftode, 2004."Traffic view: Trafficdata dissemination using car-to-car communication,"ACM SIGMOBILE Mobile Computing and Communications Review, vol. 8, no. 3, pp. 6–19,

[14] Dietzel S., Bako B., Schoch E., and Kargl F., 2009. "A fuzzy logic based approach for structure-free aggregation in vehicular ad-hoc networks,"in Proc. ACM international workshop on Vehicular Inter-networking, New York, NY.

[15] C. Sommer, O. K. Tonguz, and F. Dressler. “Traffic information systems: efficient message dissemination via adaptive beaconing,” Communications Magazine, vol. 49, no. 5, pp. 173–179, 201.

[16] Raya M.and Hubaux J.P.,2005. “The Security of Vehicular Ad-hoc Networks,” in 3rd ACM workshop on Security of Ad-hoc and sensor networks (SASN).

[17] Sapna S. Kaushik ,2013.“Review Of Different Approaches For Privacy Scheme In Vanets” International Journal Of Advances In Engineering & Technology.Vol. 5, Issue 2, pp. 356-363.

[18] John Moses S., Anitha Christy Angelin P. 2013. “Enhancing the Privacy through Pseudonymous Authentication and Conditional Communication in Vanets” Research Inventy: International Journal Of Engineering And Science Issn: 2278-4721, Vol. 2, Issue 7, Pp 45-49.www.Researchinventy.Com

[19] Jose Maria de Fuentes, Ana Isabel Gonzalez-Tablas, and Arturo Ribagorda, 2009. "Overview of Security issues in Vehicular Ad Hoc Networks", Handbook of Research on Mobility and Computing,

[20] Chadha D., Reena, Vehicular Adhoc Network (VANETs): A Review, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 3, March 2015.

[21] Zhang, Mingliu, and R. S. Wolff, ’Routing protocols for vehicular ad hoc networks in rural areas’, Communications Magazine, IEEE, 46.11, pp.19-131, 2008.

[22] Khan, Imran, and Qayyum A., 2009.‘Performance evaluation of AODV and OLSR in highly fading vehicular ad hoc network environments’,INMIC IEEE 13th International Conference.

[23] Hashimi, Haider N, Kamalrulnizam Abu B, and Kayhan Zrar Ghafoor,2011. ‘Inter-domain proxy mobile ipv6 based vehicular network’,Network Protocols and Algorithms, vol.2, issue 4, pp. 1-15.

International Journal of Computer Applications Technology and Research
Volume 5– Issue 5, 292 - 294, 2016, ISSN:- 2319–8656

Approved TPA along with Integrity Verification in Cloud

Krathika A
Department of Computer Science and Engineering
Shree Devi Institute of Technology, Kenjar, Mangalore, Karnataka
India

Abstract

Cloud computing is new model that helps cloud user to access resources in pay-as-you-go fashion. This helped the firm to reduce high capital investments in their own IT organization. Data security is one of the major issues in cloud computing environment. The cloud user stores their data on cloud storage will have no longer direct control over their data. The existing systems already supported the data integrity check without possessions of actual data file. The Data Auditing is the method of verification of the user data which is stored on cloud and is done by the TTP called as TPA. There are many drawbacks of existing techniques. First, in spite some of the recent works which supports updates on fixed-sized data blocks which are called coarse-grained updates, do not support for variable-sized block operations. Second, an essential authorization is missing between CSP, the cloud user and the TPA. The newly proposed scheme will support for Fine-grained data updates on dynamic data using RMHT algorithm and also supports for authorization of TPA**.**

Keywords: Cloud Computing, Data Security, TPA, Fine-Grained Data Updates, RMHT Algorithm

1. INTRODUCTION

A. Cloud Computing:

It is being intensively mentioned as one of the most dominant innovation in information technology in recent period. By using resource virtualization cloud will deliver us computing components and services in a pay-as-you-go scheme, by which the cloud visualize to turn into as suitable to use similar to way of life needs such as electricity, irrigate, telephone and water in the upcoming future. The various cloud computing services that are categorized into IaaS, PaaS and final one is SaaS. Many multinational IT corporations now present a powerful public cloud services to the users on a level from individual to endeavor all over the world. As we know the present growth and propagation of cloud compute is fast increasing, debate and hesitation on the practice of cloud still going.

B. Fine-Grained Data Updates:

The data owner can utilize clouds SaaS concept to store data. The data files which are stored are in the format of fixed-sized data blocks that limits operations like modification, insertions and deletions on blocks. This results in storage and data processing overheads in cloud. To solve this problem of existing systems, in proposed system the RMHT algorithm is implemented.

2. LITERATURE SURVEY

Compared to conventional systems, scalability and elasticity were the major benefit of cloud. In this paper, we will concentrate on small and frequent data updates on variable-sized data blocks. Cloud users also need to break larger datasets into minor datasets and store them on different physical servers for privacy-preserving and reliability. The major pressing issue related to cloud is data security/privacy. It is one of the most regularly raised concerns and there is a lot of progress trying to increase cloud data security/privacy with technical approaches on CSP part.

The Integrity verification for expanded data storage has attracted huge research interest. The topic on POR was first model and discovered by Jules. But, this method can only be used to static data storage. In the same year, Ateniese discovered alike scheme which he named „provable data possession". This method provided a „blockless verification". Achievement by Shacham, gave a better POR model with stateless verification. They also founded a MAC-based private verification scheme and also the first public verification scheme which was based on BLS signature.

In second method, the generation and verification of integrity proofs are alike to signing and verification of BLS signatures. It also proved the security of both the above methods and also for the PDP method by Ateniese. Later he extended his method for increased scalability, but only partial data dynamics were supported along with predefined number of challenges.

In 2009, Erway, discovered the first PDP method based on skip list that can provide full dynamic data updates. But, defaultly it did not support for public auditability and variable-sized file blocks. Wang discovered a technique which was based on BLS signature that can provide public auditing and also full data dynamics that was his latest work on public data auditing which provided dynamics support. But, this technique lacks support for fine-grained update. Later Wang added a random masking technology to ensure that the TPA cannot conclude the raw data file from a series of integrity proofs. In this technique, they also included an approach that was first discovered in to segment file blocks into multiple sectors. But in this technique, the use of this approach was restricted to trading-off storage cost with communication cost.

The author of paper PDP discovered the technique that can be used to provide dynamic operations such as deletion, modification, updating on data blocks by avoiding use of large encryptions.

3. OBJECTIVE

Contribution in this paper is listed below as follows:

1. We encourage the authorized public auditing method for ensuring secure verification of file by TPA in cloud domain. TPA cannot retrieve important information about user data. TPA is made authorized by swapping a keying material with CSP and Data owner.
2. RMHT algorithm is used in this approach. This algorithm supports for block level operations by providing variable-sized data blocks. As a result, there will be reduction in storage and computational overheads compared to other previous schemes.

4. EXISTING METHODS

The major issues in cloud computing can persist in loss of control over certain important data of cloud user, and the lack of

security for stored file. To have authorization for TPA, swapping of credentials is required such as signatures between three participating parties in cloud. The coarse-grained data updates are available in existing methods that provides update which has fixed-sized blocks. This will result in, every small periodic updates will experience higher communicational costs.

4.1 Drawbacks of Existing System:

1. The existing methods like PDP and POR support only public data integrity checks by TPA. But, it does not ensure that if TPA is authorized or not. A malicious Third Party can set DoS attack and get important information of user data by sending multiple challenge requests to cloud. This will generate additional cost on CSP.
2. Even though Coarse-grained data updates use BLS signature technique to support for small integrity proofs, it does not support for variable-sized block updates. Data updating process is difficult in existing works. Deletion of data entire size is less than block size is not permitted. If user wants to add new data in the file, then new blocks will be produced for every small insertion of data. Update of user desired block is also not provided.

5. PROPOSED SYSTEM

In this system, it shows that the TPA is made authorized by using the signature technique in which data owner after storing data on cloud will send a signature to TPA that is encrypted with owner secrete key. Later TPA owns this key to prove that he is an authorized auditor. Using this technique privacy preserving public auditing is obtained. The proposed technique brings outstanding security and is greatly efficient. The RMHT algorithm in this technique has increased storage capability of cloud. The cost of data storage is greatly reduced.

To produce the key for the file which is stored on the cloud, key generation and file pre processing methods are used. File Pre processing is carried to separate the files into block structure. This will increase the storage structure of data files. RMHT is a data structure that will arrange the file in separate block levels such that each block can have different authenticators which are connected at the root node to form metadata. Later this metadata is used for file verification purpose. When file is updated by data owner the metadata also keeps on changing. The user sends challenge request for TPA to check integrity of file on the cloud.

5.1 Advantages of Proposed System:

1. To audit for a batch of files TPA is allowed, therefore it minimizes the time of auditing task for several files at a time.
2. To execute auditing task as asked by data owner using signature scheme then the TPA is made authorized.
3. To achieve better performance and reduce extra storage, a fine-grained technique is presented in this paper

6. SYSTEM ARCHITECTURE

Fig.1 represents the three parties participating in public auditing process. And it consists of the following:

i. Cloud Service Provider (CSP)

The duty of CSP is to issue data storage and services for users. And the CSP is the main part of the cloud because it has more storage space to store and maintain data. It gives all its services as per pay-as-you-go scheme.

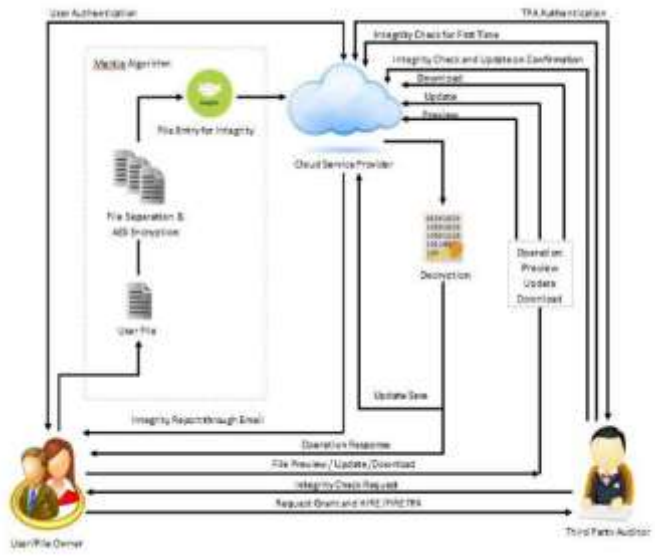


Fig. 1: System Architecture

ii. Cloud User:

A number of users can be created in which one of the users can be the group admin and can share data files with other users. And the other users can download and alter this shared file in group. If real owner of the file desires to check file for integrity then they can send challenge request to TPA. After receiving the request TPA will thus forward it to CSP and retrieves back the result of integrity proofs. Later TPA verifies proof for correctness and returns it to data owner the one who sent challenge request.

iii. Third Party Auditor (TPA):

TPA is the trusted person that store verification parameters and provide public query services for these parameters. In this paper the TTP is the one who view the user data blocks and uploads to the distributed cloud. If any alteration attempted by cloud owner a notification is sent to the TTP.

7. Scheme Used

1) Setup:

This part is based on the BLS signature technique. The client will generate keying material by using KeyGen and Fileproc. Later client will upload the data to CSS. The client will store RMHT as metadata and will authorize TPA by alloacting the value of sigAUTH.

2) Fine-grained Update Verification:

This method generates between client and CSS. The client will request for fine-grained update request to CSS via PerformUpdate and then client executes verifyUpdate algorithm to verify whether CSS has carried the update faithfully on the data block.

To update some portion of the data block, the client has to choose the PM (partial modification) process. It contains the following steps listed below:

1) In first step, the client produces an Update Request and sends to CSS. The CSS executes the PerformUpdate (UpdateRequest,F) algorithm .
2) In second step, the CSS sends the Pupdate to client then the

client executes the VerifyUpdate (Pk,; Pupdate) algorithm .

3) Challenge, Verification and show Proof generation:

In this step, TPA has to report that it is the actual one who is challenging the CSS for data integrity checking. TPA executes the GenChallenge() algorithm along with private key and signature as parameters. Later challenge message is produced with TPA‟s new ID chosen arbitrary from the set of total blocks. After this activity TPA sends challenges to CSS.

When CSS receives the challenges it will execute another algorithm to validate the signature, VID and client‟s public key. If it returns a true value, then CSS will forward a proof "p" to TPA and TPA will execute the algorithm to validate (pk, challenge, p) otherwise it returns false, and the request is rejected. For TPA permission, a signature method is chosen which cannot be faked by malicious TPAs.

The actors of three Parties Involved In Public Data Auditing Scheme is shown Using Sequence Diagram as Below:

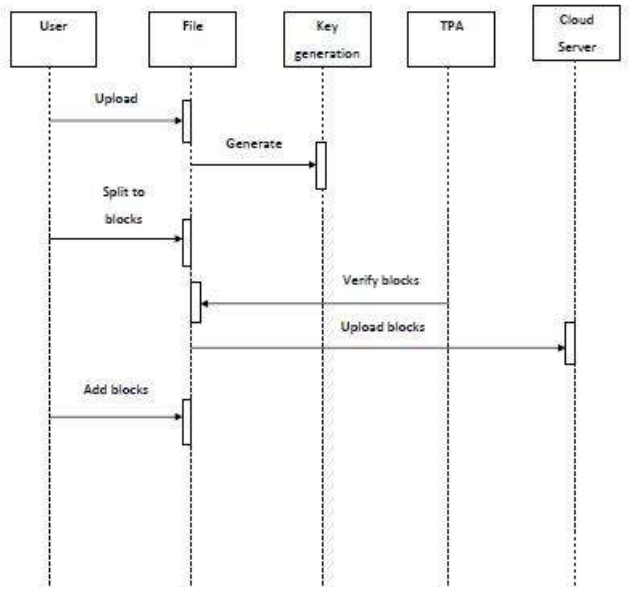


Fig. 2: Sequence Diagram

8. RESULTS AND DISCUSSIONS

Several projects progressed previously which can just store data and share data between huge numbers of user in a group. In our proposed work we have introduced a third party auditing technique to form a secure data organization process with great privacy protection technique along with working on audit ability

In this technique the main functionalities are data security, privacy protection, audit details to the data owner and finally Auditability aware data preparation

9. CONCLUSION

Cloud computing is a great computing model in which data security is the main feature for the cloud user. The newly developed technique gives support for fine-grained data updates and authorization of TPA for data security. Theoretical and experimental output for proposed system can provide higher scalability and flexibility in storing data on cloud by minimizing storage costs. This is very beneficial in big data application like social media and business transactions where minor periodic modification of data is of great importance. Security and privacy of user data is increased by using authorization scheme in public auditing. TPA cannot retrieve user data entirely during the activity of public auditing and also signature technique is used that cannot be copied so that it can avoid from malicious TPA.

10. ACKNOWLEDGEMENT

I would like to thank to my guide Prof. Rasheeda Z Khan and my parents for their highly appreciable support and encouragement .

11. REFERENCES

[1] J. Yao, S. Chen, S.Nepal,D. Levy, and J. Zic, „„TrustStore: Making Amazon S3 Trustworthy With Services Composition,"" in Proc. 10th IEEE/ACM Int"l Symposium on Cluster, Cloud and Grid Computing (CCGRID), 2010, pp. 600-605.

[2] Q. Wang, C.Wang, K. Ren,W. Lou, and J. Li, „„Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing,"" IEEE Trans. Parallel Distrib. Syst., vol. 22, no. 5, pp. 847-859, May 2011.

[3] C. Wang, Q. Wang, K. Ren, and W. Lou, „„Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing,"" in Proc. 30st IEEE Conf. on Comput. and Commun. (INFOCOM), 2010, pp. 1-9.

[4] G. Ateniese, R.D. Pietro, L.V. Mancini, and G. Tsudik, „„Scalable and Efficient Provable Data Possession,"" in Proc. 4th Int"l Conf. Security and Privacy in Commun. Netw. (SecureComm), 2008, pp. 1-10.

[5] G. Ateniese, R. Burns, R. Curtmola, J. Herring, O. Khan, L. Kissner, Z. Peterson, and D. Song, „„Remote Data Checking Using Provable Data Possession,"" ACM Trans. Inf. Syst. Security, vol. 14, no. 1, May 2011, Article 12.

[6] G.Ateniese, R.B. Johns,R. Curtmola, J.Herring, L. Kissner,Z. Peterson, and D. Song, „„Provable Data Possession at Untrusted Stores,"" in Proc. 14th ACM Conf. on Comput. and Commun. Security (CCS), 2007, pp. 598-609.

International Journal of Computer Applications Technology and Research
Volume 5– Issue 5, 295 - 303, 2016, ISSN:- 2319–8656

A Survey on Video Watermarking Technologies based on Copyright Protection and Authentication

Hedayath Basha Shaik
Assistant Professor, ECE Dept.
R.M.K. College of Engineering and Technology
Chennai, India

Gangatharan. N
Professor and Head of the Department ECE
R.M.K. College of Engineering and Technology
Chennai, India

Tamilchelvan. R
CEO
Vidhara Secure Com Company
Chennai, India

Abstract: Digital Watermark is class of marker or symbol secretly embedded in a multimedia signal such as Audio, Image or Video. It is used to identify the ownership of the multimedia signal. Video watermarking is an emerging area for various applications like copy control broadcast monitoring, video authentication, copyright protection and enhanced video coding. The main objective of this paper is to present survey and comparisons of various available techniques on video watermarking based on copyright protection and identification. Comparative study of various technologies gives the significant information about the PSNR, payload, quality factor and also the various attacks used in video watermarking techniques. The best techniques in various scenarios are discussed in this paper which will help the research scholars in field of video watermarking.

Keywords: authentication; copyright protection; video attacks; psnr; payload.

1. INTRODUCTION

Broadly the digital data is managed using the digital rights management (DRM) technologies where the DRM systems possess the following techniques they are a) encryption b) digital Certificates c) watermarking d) access control e) secure communication protocols f) fingerprinting g) rights specification language h) trust infrastructure and i) hashing. So the watermarking technique is one of the subset of DRM techniques.

Due to the enormous growth of digital multimedia technologies the use of digital signals incredibly increases, so the attention to the field of digital authentication is increasing. Digital watermarking is documented as an efficient measure for copyright protection of digital multimedia signals. The digital watermarking is classified in terms of three categories they are a) spatial domain watermarking b) frequency domain watermarking and c) feature domain watermarking. In spatial domain watermarking techniques the watermarking is done directly modifying the pixel values of the host multimedia signal. One of the simplest techniques is least significant bit (LSB) modification, where the human visual system (HVS) cannot be able to detect the changes in the original host signal. In transform domain the spatial domain host multimedia signal is transformed into frequency domain using discrete cosine transform (DCT), discrete wavelet transform (DWT) and other available transforms as required by the user. The watermarking is done on the transformed signal where the HVS cannot be able to detect the watermark and it will be completely invisible. In feature domain the watermarking is done on any one of the features of the host multimedia signal. The host multimedia feature is chosen by the user to embed the watermark where the features or entire host signal will be processed by filters like high pass filters. By selecting the edges of the host signal or the feature the embedding process will be carried out for the better authentication and copyright protection.

2. WATERMARKING SYSTEM

Audio, image and video watermarking methods uses the exclusive generic building blocks, they are a) watermark embedding system and b) watermark recovery system. The below figure. 1 and figure. 2 shows the watermarking embedding and recovery systems.

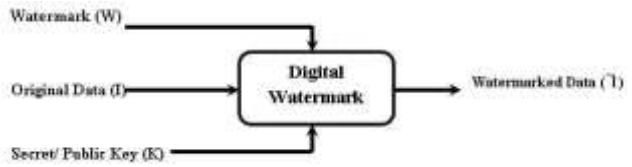


Figure 1 Generic digital watermarking system

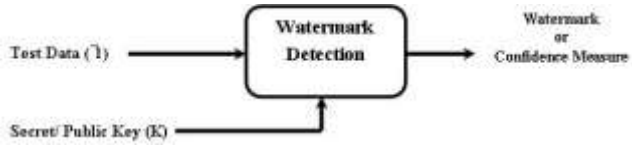


Figure 2 Generic digital watermark recovery system

Here, 'W' is the watermark used to embed in the original data (I), 'K' is the public or private key defined by the user and Ĩ is the watermarked data.

Table 1 shows the information regarding different multimedia signals, watermarking types and three different domains. In which frequency domain is mostly used where the user can able to meet certain required criterions.

Table 2 gives the information about various watermarking applications and the various requirements for the user. There is a tradeoff between the robustness, capacity and imperceptibility which is shown in the below figure 3. The watermarked data may likely to undergo either intentional or

unintentional modifications two groups of distortion can be distinguished. The first one contains distortions which can be considered as additive noise to the data whereas the distortion in the second group are due to modifications of the spatial or temporal data geometry with the intent to introduce a mismatch between the watermark and the key used for embedding.

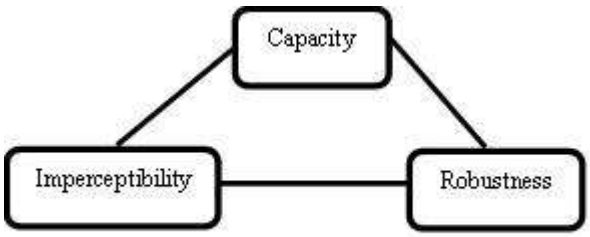


Figure 3 Tradeoff triangle in watermarking system

In the video watermarking process after watermark embedding the watermarked video is send through the channel where the video can under go various attacks.

The watermarked data may likely to undergo either intentional or unintentional modifications two groups of distortion can be distinguished. The first one contains distortions which can be considered as additive noise to the data whereas the distortion in the second group are due to modifications of the spatial or temporal data geometry with the intent to introduce a mismatch between the watermark and the key used for embedding [2]

Table 1. Watermarking domains and types

Multimedia Signals used for watermarking	Watermarking Domains	Types of Watermarking
Text Documents or Audio Files or Image Files or Video Signals	a. Spatial Domain Watermarking b. Frequency domain Watermarking c. Feature domain Watermarking	a. Private Watermarking (also called Non – Blind Watermarking) b. Semiprivate Watermarking (also called Semi-Blind Watermarking) c. Public Watermarking (also called as Blind or Oblivious Watermarking)

Table 2. Watermarking applications and requirement

Watermarking Applications	Watermarking Requirements
Watermarking for Copyright Protection	Recovery with or without the original data.
Fingerprinting for Traitor Tracking	Extraction or Verification of a given watermark
Watermarking for Copy Protection	Robustness
	Imperceptibility
Watermarking for Image and Video Authentication	Security issues and use of keys

3. VIDEO WATERMARKING SYSTEMS

Video is referred as moving still images or moving frames, video watermarking is done by watermarking each still image using the image watermarking algorithms. It is obviously known that video data is massive in size, for an example if a digital video where each frame is of 720 X 480 pixels with color data i.e. (24 bits per pixel) then 30 frames/ second has a rate of 248 Mbps. Due to the high rate extension of image watermarking techniques and architectures may be inefficient to handle video watermarking. This is one of the reason by which researchers got motivated to design low complex and efficient hardware implementation [3].

Digital video can be copied frequently exclusive of quality loss. Thereby copyright protection of video is one of the significant issues in digital video transmitting networks than it was with analog TV broadcast. Frank Hartung and Bernd Girod proposed robust watermarking of MPEG 2 encoded video. This scheme is possessing lower complexity comparing with decoding process with watermarking and re-encoding. In this scheme 'C' Programming is used to take input and the DCT is applied to the parts of bit stream which contains DC and AC coefficients and are replaced by DC and AC coefficients with watermark. This method is very robust against un-attempted and attempted attacks. It can be applied to MPEG -1, ITU – T H.261 or ITU – T H.263 video coding schemes; the embedded watermark is robust against linear and nonlinear operations like cropping, filtering and quantization in pixel or frequency domain [4].

Chiou – Ting Hsu and Ja – Ling Wu proposes DCT based watermarking for video to hide covert information into signals to protect the authentication and copyright of the digital MPEG videos. Here the watermark is embedded in intra and non-intra frame with dissimilar residual masks. This method is robust to cropping operation and MPEG compression [5].
Frank Hartung and Bernd Girod propose additive spread spectrum methods for embedding watermarks into un-compressed video and compressed MPEG2 video. It is important practically to work on encoded rather than un-encoded video. Here the watermark is embedded in entropy coded DCT coefficients [6].

Mitchell D. Swanson et. al. presented a video watermarking method to enable copyright protection into a digital video in which it is based on multi resolution scene based and video dependent watermarking. This method provides imperceptible watermarking and there are two pseudorandom keys used. The watermarking procedure is very robust to several distortions and degradation [7].
Tae – Yun Chung et. al., proposes video watermarking technique by extending the direct spread spectrum on MPEG2

video. In their proposed technique they increases the video perception quality of the embedded watermarked video by controlling the parameters like strength and area of embedding with respect to the characteristics of the video. The average PSNR value for the three different experimental videos is 29.5 dB for 4Mbps, 30.6dB for 5Mbps, 31.6dB for 6Mbps and 32.1dB for 7Mbps, the watermark detection rate is above 97% for every bit rated MPEG2 video signal [8].

Christoph Busch et. al., proposes modified Koch – Zhao algorithm for video watermarking and observing video streams in a TV broadcasting environment which survives attacks like MPEG2 video compression. The algorithm is well suited for digital video watermarking video streams such as sporting events or movies [9].

Wenwu Zhu et. al., presents a unique approach for Images and video watermarking, their approach is based on 2D and 3D DWT. For images SPIHT algorithm is used for compression and the PSNR values between watermarked and original is about 42.77dB. In concerned with video the watermark is embedded in Group of Pictures (GOP) 16 frames of QCIF foreman sequence 3D wavelet based SPIHT video coder is used for compressing video, the PSNR of original and watermarked frames is 29.99dB and 29.46dB [10].

Gerhard C. Langelaar et. al., gives a overview of digital and video data watermarking, concerned with video data DCT transform with pseudo random noise is discussed and it is stated that the "Robustness of a watermark is improved by improving the energy of the watermark" also stated that " In the real time environment computational complexity and robustness play very significant role" [11].

Xiamu Niu et. al., team projected a multi-resolution watermarking technique where a gray scale image is embedded into a digital video, 2D and 3D of the video signals are used for embedding watermark. The Hamming code, 2D and 3D Discrete Wavelet Transform (DWT) is used for the signal processing. This watermark is robust against attacks like lossy compression, averaging and frame dropping [12].

T. Brandao et. al., team proposes an analysis on the effects of signal mixture techniques in video watermark detection. Spread spectrum video watermarking is used and various common error correction codes such as BCH, Reed – Solomon, with multilevel signaling, Binary Convolution Codes (BCC) with Viterbi decoding is done to improve results [13].

Emmanuel Garcia et. al., introduced a novel framework on texture based watermarking of 3D video objects. To main objective is to obtain the information hidden in the texture image without degrading the visual perception. Here blind watermarking is proposed using EUREMARK algorithm [14]. Shih – Wel Sun and Pao – Chi Chang presents a new approach based on temporal synchronization. Video watermarking is done through matching the profile statistics. It is given that by position mean and variance in the X and Y directions of the frame is send to the receiver to check the received data. The accuracy of frame detection is from 72.41% to 98.15% [15].

Gwenael Doerr and Jean – Luc Dugelay done four approaches in the video watermarking based on spread spectrum (SS) technique.

a) Every frame is embedded with different watermark with embedding strength parameter 'α', and secret key 'K'. (Uncorrelated watermark embedding) (SS system), Wt(K) is the inserted watermark has normal distribution with zero mean and unit varience.
b) Only one watermark is used to embed into the video frames (Redundant watermark embedding) (SS-1 System)
c) Embedding different watermarks randomly in the video frames (SS-N)
d) SS-α

Previous works have mainly focused on robustness i.e. resilience against non-malicious attacks. For example, for applications such as broadcast monitoring, video authentication or data hiding, the watermark has to undergo some signal processing e.g. noise addition, filtering, lossy compression. However, for fingerprinting or copy-control applications, the embedded watermark has also to survive in a hostile environment with malicious users. In this context, security issues have to be addressed [16].

Eugene T. Lin et. al., developed a state machine key generator which help the user to detect the watermark even if synchronization of the video signal is lost. This blind watermarking technology is used to establish and maintain temporal synchronization. It is very resilient against temporal synchronization attacks [17].

Mauro Barni et. al., team members confidently proposed a method to watermark MPEG-4 video objects in a very efficient manner. The proposed method embeds the watermark in each video object by posing a particular relationship between some predefined pairs of quantized DCT coefficients in the luminance block. Watermark is equally embedded into Inter and Intra macro blocks of the video [18] . Satyen Biswas et. al., team works on uncompressed video and also on the compressed video sequences. The main theme is to improve the authentication of multimedia objects and it is done using watermarking in the GOP. Drift Compensation method is used to predict the changes between successive frames in the video clips. Blind detection of watermark is done using estimation technique and it is very robust to attacks [19].

Karen Su et. al., presents a hypothetical structure for the linear collusion analysis of digital video watermarking, derive new statistical invisibility theorem, collusion – resistance theorem and practical design rules. Here the design of a copyright protection system for MPEG2 videos and attacks of multiple frames linear collusion like a) Linear Collusion, b) Statistical Invisibility are also discusses [20].

Yulin Wang and Alan Pearmain, presents MPEG2 video blind watermarking and this is very robust to geometric attaks. The proposed method is not restricted to MPEG2 alone it can be suited for DCT based coding videos [21].

Jing Zhang et. al., presents a robust video watermarking of H.264/AVC where pre – processing of the gray scale watermark is done to obtain the 1D output sequence. This obtained pattern is embedded into a compressed video. It achieves high robustness to various signal processing attacks and good visual quality [22].

Maneli Noorkami and Russell M. Mersereau, introduced a structure of robust video watermarking of H.264 video to

provide copyright protection and authentication. In this work the authors uses 4X4 DCT to increase the payload of the watermark and also uses a key dependent algorithm to have a visual watermarking capacity. This work provides good robustness to various attacks [23].

Alper Loz and Aydin Alatan, compares their work named spatio – temporal watermarking of video by using HVS with the other two algorithms i.e. first one is based on spatial sensitivity of HVS and the other uses only the HVS characteristics. It is present that the robustness of the watermark can be improved by integrating temporal characteristics. This work is robust against common temporal signal processing operations [24].

Siyue Chen and Henry Leung, presents disordered semi fragile watermarking for video authentication used in scrutiny applications, here they used raw video data for processing. Mapping is carried on GOP and frame index separately. This method is robust to common spatial processing [25].

Lino E. Coria et. al., discourage camcorder in theater to avoid the piracy of the videos. The team developed a system using dual tree complex wavelet transform for watermarking, the video watermarking is performed such that if the watermark is displayed the video will not be played in a player it is very robust to geometric distortions and lossy compressions [26].

Young-Yoon Lee et. al., presents two different temporal feature modulation algorithms where in the first one is the watermark is embedded in skipping selected frames and the second is to find the centers of gravity in the blocks to embed the watermark. It is robust to compression and temporal attacks [27].

R. Reyes et. al., proposes a video watermarking system where the watermark is logo image of the owner. In this work the video sequences are segmented in every frame and the watermark logo is embedded into the frames randomly. The security is increased by using a logo binary pattern mapped to a noise like binary pattern before embedding this process is robust against several attacks [28].

Alper Koz et. al., team presents watermarking of free view video in which the watermark is embedded with three different factors where the first one global scaling factor, second one is image processed through high pass filter and the last watermark sequence with zero mean and unit variance. The watermark is embedded in every frame of the video and it is robust to geometric and compression attacks [29].

Min – Jeong et. al., presents a very good approach to avoid the pirates copying the digital cinema using camcorder, this technique provides robust watermark recognition in opposition to camcorder capture and also to extract data about the place and time of piracy. It helps to find out the persons performing pirates by using position estimate model approach [30].

Liyun Wang et. al., proposes real time video watermarking scheme on compressed videos like MPEG1, MPEG2 and it can be applied for MPEG4 and H.264 because DWT domain could directly acquired from block DCT's of any size. This work is very robust to geometric attacks and also used for hiding of data [31].

Andras Boho et. al., presents video security and tackles cryptography and signal processing operations each other. It uses little encryption techniques in tradeoff between preserved functionality and security. The encryption of data sets in H.264/AVC and HEVC achieves consistently low SSIM values. The watermark is robust against signal processing operations and transcoding. It achieves better tradeoff between robustness, perceptibility and payload [32].

Mehdi Fallahpour et. al., presents a sensible system of digital video watermarking for tamper detection of compressed videos and authentication. The embedding is done in LNZ DCT blocks and extracting of watermarks are integrated with the coding and decoding of video codec. The cryptography technique is used to improve the security of the system [33].

Md. Asikuzzaman et. al., team presents three different versions of robust blind video watermarking based on DT – CWT. First the watermark is embedded in the level of three coefficient of a three level DT – CWT decomposition of chrominance channel to make robust to geometric attacks [34].

In Table 3 and 4 the detailed information of the existing video watermarking techniques in terms of videos been used, watermarking domains, attacks and experimental results.

4. STATISTICS OF VIDEO WATERMARKING SYSTEM

The below figures 4, 5 and 6 shows the statistics of the video watermarking systems been used in research where the compressed videos utilization is more compared with the uncompressed video or raw videos. The main reason is when the raw video is processed with watermark embedding process the data is getting effected and the quality decreases as such psnr value is decreased. In figure 5 the type of domain usage statistics is performed and in figure 6 psnr of various systems are discussed.

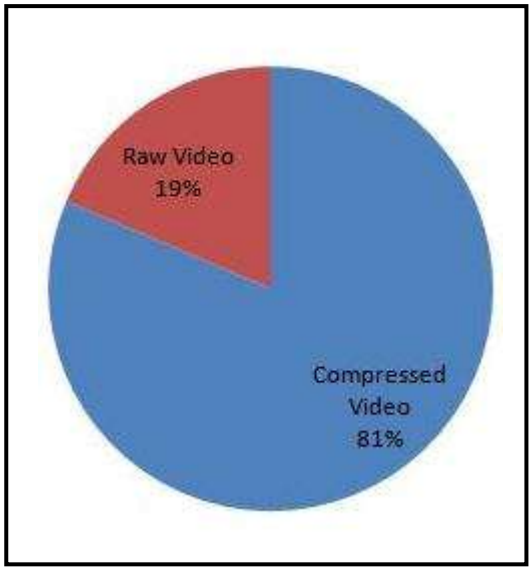


Figure 4 Percentage of Video formats used

Table 3. Detailed video watermarking domains

S.No	Authors	Type of video used	Video Watermarking Domain
1	Frank Hartung and Bernd Girod [4].	Compressed MPEG2 Video	Spatial and Frequency Domain: DCT
2	Chiou – Ting Hsu and Ja – Ling Wu [5]	Compressed MPEG Video	Frequency domain: DCT
3	Frank Hartung and Bernd Girod [6]	Raw video and Compressed MPEG2 Video	Spatial and Frequency domain:DCT
4	Mitchell D. Swanson et. al. [7]	Compressed Video data	Spatial and Frequency domain
5	Tae – Yun Chung et. al.,[8]	Compressed MPEG2 Video	Frequency domain: DCT
6	Christoph Busch et. al.,[9].	Compressed MPEG2 Video	Spatial and Frequency domain: Modified Koch – Zhao Algorithm (New DCT & IDCT Algorithms)
7	Wenwu Zhu et. al.,[10]	JPEG Image and MPEG2 Video	Frequency Domain: DWT
8	Xiamu Niu et. al., [12]	Compressed MPEG Video	Frequency domain: DWT
9	T. Brandao et. al., [13]	Compressed MPEG2 Video	Spatial domain: Spread Spectrum
10	Emmanuel Garcia et. al. [14].	3D Objects	Spatial Domain: Pseudorandom noise is added as watermark
11	Shih – Wel Sun and Pao – Chi Chang [15]	Uncompressed Video	Spatial Domain
12	Gwenael Doerr and Jean – Luc Dugelay, [16].	MPEG2 Compressed Video	Spatial Domain
13	Eugene T. Lin et. al.,[17]	Uncompressed Video	Spatial Domain
14	Mauro Barni et. al. [18]	Compressed MPEG-4 video	Transform domain
15	Satyen Biswas et. al., [19]	Uncompressed (DAVI) and Compressed MPEG2 Video	Spatial domain & Transform domain (DCT)
16	Yulin Wang and Alan Pearmain, [21]	Compressed MPEG2 video	Transform domain: DCT
17	Jing Zhang et. al. [22]	Compressed Video H.264/AVC	Transform domain: DCT
18	Maneli Noorkami and Russell M. Mersereau, [23]	Compressed Video H.264	Transform Domain: DCT
19	Alper Loz and Aydin Alatan, [24]	Compressed video ITU H.263	Spatial and Transform domain: DCT
20	Siyue Chen and Henry Leung, [25]	Raw video (AVI)	Transform domain: DCT
21	Lino E. Coria et. al., [26]	Standard Video files of QCIF (176 X 144)	Transform domain DT-CWT and DWT.
22	Young-Yoon Lee et. al,[27]	Compressed Video H.264/AVC	Spatial domain
23	R. Reyes et. al.,[28]	Compressed Video	Transform domain:DWT
24	Alper Koz et. al.,[29]	TV Video Signal	Spatial domain
25	Min – Jeong et. al., [30]	Watermarking the Pirates camcorder	Spatial Domain
26	Liyun Wang et. al.,[31]	MPEG1 and MPEG2 Videos	Frequency domain: DCT & DWT
27	Andras Boho et. al., [32]	H.264/AVC and HEVC video formats	Spatial and Frequency domains
28	Mehdi Fallahpour et. al., [33]	H.264/AVC video formats	Spatial and Frequency domain
29	Md. Asikuzzaman et. al.,[34]	MPEG video formats	Frequency domain: DT CWT

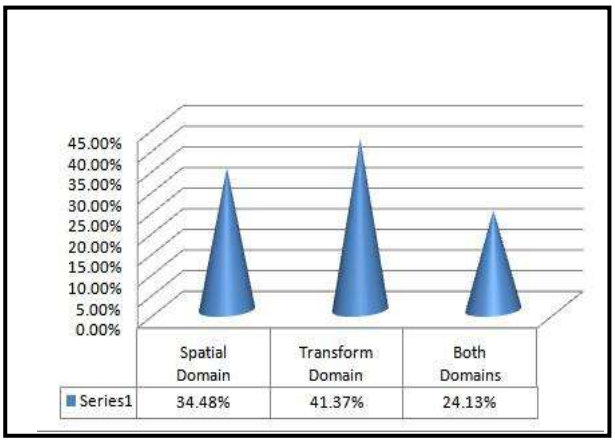


Figure 5 Percentage of used Domains

Table 4 Video Watermarking Attacks and Results

S.No	Authors	Attacks performed	Experimental Results
1	Frank Hartung and Bernd Girod [4].	Robust against Un-attempted and Attempted attacks	15 to 30% of DCT coefficients are altered.
2	Chiou – Ting Hsu and Ja – Ling Wu [5]	Robust to Cropping and MPEG Compression	PSNR is 44.63dB NC=1
3	Frank Hartung and Bernd Girod [6]	Robust to temporal attacks and MPEG compression	Chip Rate is 633,600 Error is 0.0194
4	Mitchell D. Swanson et. al. [7]	Robust to several Video degradation and distortion.	Avg. PSNR is 36.9dB
5	Tae – Yun Chung et. al.,[8]	Robust to Cropping and MPEG Compression	Watermark Extraction Accuracy is 97% Avg. PSNR is 32.5
6	Christoph Busch et. al.,[9].	Robust against MPEG2 encoding Restricted to some well of geometric attacks	Transparent Watermarking is attained Data Rates are 4Mbps and 6Mbps Percentage of Corrected bits is above 90%
7	Wenwu Zhu et. al.,[10]	Robust to Image/ Video compression and digital halftoning.	PSNR(Image) is 42.77dB PSNR(Image) after SPIHT compression is 33.48dB PSNR of Video first frame is 29.99dB Computational saving is 87.5%
8	Xiamu Niu et. al., [12]	Robust to Frame dropping, averaging and lossy compression	PSNR values of original and waternarked frames of size 352X240 is 32.49
9	T. Brandao et. al., [13]	Robust to compression attack	Presents a good Performance vs. simplicity trade off.
10	Emmanuel Garcia et. al. [14].	Robust to geometric disturbances	Better tradeoff between capacity and robustness is achieved. Blind Watermarking is done
11	Shih – Wel Sun and Pao – Chi Chang [15]	Robust to temporal attacks like transposition, dropping and insertion.	The accuracy of frame detection is from72.41% to 98.15%.
12	Gwenael Doerr and Jean – Luc Dugelay, [16].	Robustness is achieved for every attack step by step	Four modules are developed in spread spectrum
13	Eugene T. Lin et. al.,[17]	Robust against temporal Synchronization attacks	Developed a state machine key generator. Flicker is generated in the video due to watermark
14	Mauro Barni et. al. [18]	Robust against Bit rate decreasing, Frame dropping,	Synchronization is achieved. Confidence values are high when wrong key is used.
15	Satyen Biswas et. al., [19]	Robust against Spatial and temporal Attacks	Average NC = 0.95 and decreases with increase in frame dropping.
16	Yulin Wang and Alan Pearmain, [21]	Robust to geometric attacks	Q Step is 6, 8 and 12. Error rate is below 1.7%
17	Jing Zhang et. al. [22]	Robust to Transcoding and Signal processing Attacks like Gaussian low pass filtering Additive Gaussian Noise(Variance is 0.75) Circular averaging filter Unsharp Contrast enhancement	Correlation values for watermarked video is 0.93, Correlation after transcoding (1/3rd Bit rate) is 0.50 Correlation after Circular filtering is 0.87 Correlation after Contrast enhancement is 083 Correlation after Gaussian Noise is 0.75
18	Maneli Noorkami and Russell M. Mersereau, [23]	Robust to Several common signal processing attacks	High Payload Average Correlation factor is 0.997
19	Alper Loz and Aydin Alatan, [24]	Robust to Several common signal processing attacks Transcoding attacks	Average PSNR(Foreman) is 39dB Avg. PSNR (Mother Sequence) is 43dB.

20	Siyue Chen and Henry Leung, [25]	Robust to JPEG Compression Median Filtering Contrast Enhancement	Data Payloads are 32 and 1024 bits PSNR is 42dB
21	Lino E. Coria et. al., [26]	Robust to Attacks Fragile to Joint Attack	Avg. PSNR is 41dB
22	Young-Yoon Lee et. al,[27]	Robust against Compression and temporal attacks	Frames rate is 29.977 or 23.976 fps. PSNR is 31dB
23	R. Reyes et. al.,[28]	Robust against different Noisy attacks	NC= 0.95
24	Alper Koz et. al.,[29]	Robust to AWGN noise compression Attacks Geometric Attacks	Best NC=0.875 Avg. Noise PSNR is 42.11
25	Min – Jeong et. al., [30]	Robust to camcorder capture Geometric distortions Signal Processing distortions	Avg PSNR is 46.0 dB Avg PSNR is 45dB for HD videos
26	Liyun Wang et. al.,[31]	Robust against cropping and rotation attacks	Avg. Bit Error Rate (BER) is almost negligible i.e. Zero for many attacks. Avg. BER is 3.3 for H.264 Compression
27	Andras Boho et. al., [32]	Robust against Signal Processing attacks Transcoding Compression	Quantization Index Modulation is employed Avg. H.264/AVC PSNR is 55dB Avg. BER is 0.011
28	Mehdi Fallahpour et. al., [33]	Robust to common video processing operations	PSNR is 50.54 PSNR degradation is 0.88dB Structural Similarity Index Decreases is 0.0090 Bit Correct rate is 0.71 to 0.88
29	Md. Asikuzzaman et. al.,[34]	Robust to Compression and Geometric Attacks Temporal Synchronization Attacks	SSLD, SDLD and KDLD methods were proposed

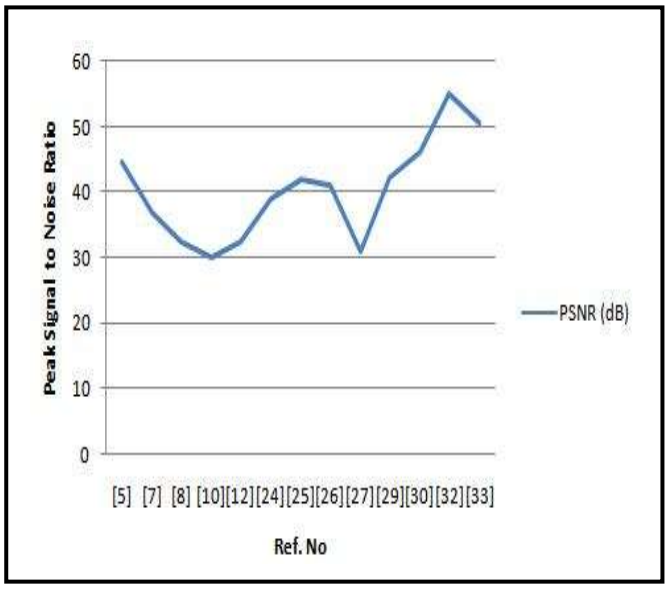


Figure 6 PSNR values of Video watermarking systems

5. CONCLUSION

In this video watermarking survey [4] to [34] are using compressed video except [14], [15], [17] and [25] where these uses un-compressed video or raw video for video watermarking. In [3] and [19] video watermarking is for both Compressed and Uncompressed video formats. The best technique for the compressed video watermarking in terms of PSNR can be attained by using parametric lattice quantization index modulation (QIM) proposed by Andras Boho et. al., [32] and also proved that Structural Similarity (SSIM) index is considered as better parameter than that of PSNR to calculate the quality degradation [32]. In the attempt to provide robustness against various video attacks Keyless Dynamic Level Detection (KDLD) method proposed by Md. Asikuzzaman et. al.,[34] provides best robustness to the various attacks compared with other methods. Min – Jeong et. al., [30] gives the best position estimating model (PEM) with a mean absolute error (MAE) to prevent pirated copies of digital cinema captured by tripod mounted SONY HDR FX 1 camcorder in real theaters. The best technique for uncompressed video is chaotic semi fragile watermarking which gives PSNR of 42dB proposed by Siyue Chen and Henry Leung [25].

6. REFERENCES

[1] Gwenael Doerr and Jean-Luc, "Security Pitfalls of Frame by Frame Approaches to Video Watermarking", IEEE Transactions on Signal Processing, Supplement on secure media, October 2004.

[2] Stefan Katzenbeisser and Fabien A.P. Petitcolas, "Computer Security Series: Information Hiding Techniques for Steganography and Digital Watermarking", Artech House Boston, London.

[3] Ali Mohammad Al – Haj, "Premier Reference Source – Advanced Techniques in Multimedia Watermarking: Image, Video and Audio Applications" Princess Sumaya University for Technology, Jordan.

[4] Frank Hartung and Bernd Girod, "Copyright Protection in Video Delivery Networks by Watermarking of Pre – Compressed Video", Multimedia Applications, Services and Techniques – ECMAST' 97, Springer Lecturer Notes in Computer Science, Vol. 1242, pp. 423 – 436.

[5] Chiou – Ting Hsu and Ja – Ling Wu, "DCT based Watermarking for Video", IEEE Transactions on Consumer Electronics, Vol.44, No.1, Feb 1998, pp. 206 – 216.

[6] Frank Hartung and Bernd Girod, "Watermarking of Uncompressed and Compressed Video", Elsevier Preprint, Vol. 66, No. 3, May 1998, pp. 283 – 301.

[7] Mitchell D. Swanson et. al. "Multiresolution Scene – Based Video Watermarking using Perceptual Models", IEEE Journal on Selected areas in Communications, Vol. 16, No. 4, May 1998, pp. 540 – 550.

[8] Tae – Yun Chung et. al., "Digital Watermarking for Copyright Protection of MPEG2 Compressed Video", IEEE Transactions on Consumer Electronics, Vol. 44, No. 3, August 1998, pp. 895 – 901.

[9] Christoph Busch et. al., "Digital Watermarking: From Concepts to Real time Video Applications", Image Security, IEEE Computer Graphics and Applications, January/February 1999, pp. 25 – 35.

[10] Wenwu Zhu et. al., "Multiresolution Watermarking for Images and Video", IEEE, Transactions on Circuits and Systems for Video Technology, Vol. 9, No. 4, June 1999, pp. 545 – 550.

[11] Gerhard C. Langelaar et. al., "A State of the Art Overview: Watermarking Digital Image and Video Data", IEEE Signal Processing Magazine, September 2000, pp. 20 – 46.

[12] Xiamu Niu et. al., "Multiresoultion Watermarking for Video based on Gray Level Digital Watermark", IEEE Transaction on Customer Electronics, Vol. 46, No. 2, May 2000, pp. 375 – 384.

[13] T. Brandao et. al., "Diversity Enhancement of Coded Spread Spectrum Video Watermarking", Wireless Personal Communications, Kulwer Academic Publishers, 2002, pp. 93 – 104.

[14] Emmanuel Garcia et. al., "Texture Based Watermarking of 3D Video Objects", IEEE transaction on Circuits and Systems for Video Technology, Vol. 13, No. 8, August 2003, pp. 853 – 866.

[15] Shih – Wel Sun and Pao – Chi Chang, "Video Watermarking Synchronization based on Profile Statistics", IEEE A&E Systems Magazine, May 2004, pp. 21 – 25.

[16] Gwenael Doerr and Jean – Luc Dugelay, "Security Pitfalls of Frame by Frame Approaches to Video Watermarking", IEEE Transactions on Signal Processing, Vol. 52, No. 10, October 2004, pp. 2955 – 2964.

[17] Eugene T. Lin et. al., "Temporal Synchronization in Video Watermarking", IEEE Transactions on Signal Processing, Vol. 52, No. 10, October 2004, pp. 3007 – 3022.

[18] Mauro Barni et. al., "Watermarking of MPEG-4 Video Objects", IEEE Transactions on multimedia, Vol. 7, No. 1, February 2005, pp. 23 – 32.

[19] Satyen Biswas et. al., "An Adaptive Compressed MPEG-2 Video Watermarking Scheme", IEEE Transactions on Instrumentation and Measurement, Vol. 54, No. 5, October 2005, pp. 1853 – 1861.

[20] Karen Su et. al., "Statistical Invisibility for Collusion – Resistant Digital Video Watermarking", IEEE Transactions on Multimedia, Vol. 7, No. 1, February 2005, pp 43 – 51.

[21] Yulin Wang and Alan Pearmain, "Blind MPEG2 Video Watermarking Robust Against Geometric Attacks: A set of Approaches in DCT domain", IEEE Transactions on Image Processing, Vol. 15, No. 6, June 2006, pp. 1536 – 1543.

[22] Jing Zhang et. al.,"Robust Video Watermarking of H.264/AVC", IEEE Transactions on Circuits and Systems-II:Express Briefs, Vol. 54, No. 2, February 2007, pp. 205 – 209.

[23] Maneli Noorkami and Russell M. Mersereau, "A Framework for Robust Watermarking of H.264 – Encoded Video with Controllable Detection Performance", IEEE Transactions on Information Forensics and Security, Vol. 2, No. 1, March 2007, pp. 14 – 23.

[24] Alper Loz and Aydin Alatan, "Oblivious Spatio – Temporal Watermarking of Digital Video by Exploiting the Human Visual System", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 18, No. 3, March 2008, pp. 326 – 337.

[25] Siyue Chen and Henry Leung, "Chaotic Watermarking for Video Authentication in Surveillance Applications", IEEE Transactions on Circuits and Systems for Video Technology", Vol. 18, No. 5, May 2008, pp. 704 – 709.

[26] Lino E. Coria et. al., "A Video Watermarking Scheme based on the Dual Tree Complex Wavelet Transform", IEEE Transactions on Information Forensics and Security, Vol. 3, No. 3, September 2008, pp. 466 – 474.

[27] Young-Yoon Lee et. al, "Temporal Feature Modulation for Video Watermarking", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 19, No. 4, April 2009, pp. 603 – 608.

[28] R. Reyes et. al., "Digital Video Watermarking in DWT Domain using Chaotic Mixtures", IEEE Latin America Transactions, Vol. 8, No. 3, June 2010, pp. 304 – 310.

[29] Alper Koz et. al., "Watermarking of Free – View Video", IEEE Transactions on Image Processing, Vol. 19, No. 7, July 2010, pp. 1785 – 1797.

[30] Min – Jeong et. al., "Digital Cinema Watermarking for Estimating the Position of the Pirate", IEEE Transactions on Multimedia, Vol. 12, No. 7, November 2010, pp. 605 – 621.

[31] Liyun Wang et. al., "Real Time Compressed Domain Video Watermarking Resistance to Geometric Distortions", IEEE Multimedia in Forensics, Security and Intelligence, Computer society, January – March 2012, pp. 70 – 79.

[32] Andras Boho et. al., "End to End Security for Video Distribution", IEEE Signal Processing Magazine, March 2013, pp. 97 – 107.

[33] Mehdi Fallahpour et. al., "Tampering Detection in Compressed Digital Video Using Watermarking" IEEE Transactions on Instrumentation and Measurement, Vol. 63, No. 5, May 2014, pp. 1057 – 1072.

[34] Md. Asikuzzaman et. al., "Imperceptible and Robust Blind Video Watermarking using Chrominance Embedding: A set of Approaches in the DT CWT Domain" IEEE Transactions on Information Forensics and Security, Vol. 9, No. 9, September 2014, pp. 1502 – 1517.

International Journal of Computer Applications Technology and Research
Volume 5– Issue 5, 304 - 311, 2016, ISSN:- 2319–8656

A Proactive Approach in Network Forensic Investigation Process

Joseph MbuguaChahira
GarissaUniversityCollege,
Garissa- Kenya

Jane KinanuKiruki,Chuka University,
Chuka- Kenya

Peter KipronoKemei
Egerton University,
Nakuru- Kenya

Abstract

Information Assurance and Security (IAS) is a crucial component in the corporate environment to ensure that the secrecy of sensitive data is protected, the integrity of important data is not violated, and the availability of critical systems is guaranteed. The advancement of Information communication and technology into a new era and domain such as mobility and Internet of Things, its ever growing user's base and sophisticated cyber-attacks forces the organizations to deploy automated and robust defense mechanism to manage resultant digital security incidences in real time. Digital forensic is a scientific process that facilitates detection of illegal activities and in-appropriate behaviors using scientific tools, techniques and investigation frameworks. This research aims at identifying processes that facilitate and improves digital forensic investigation process. Existing digital forensic framework will be reviewed and the analysis will be compiled toderive a network forensic investigation framework that include evidence collection, preservation and analysis at a sensor level and in real time. It is aimed to discover complete relationship with optimal performance among known and unseen/new alerts generated by multiple network sensors in order to improve the quality of alert and recognize attack strategy

Key words: Digital forensic, cybercrimes, proactive network forensic, attack prediction, attack Strategy.

1.0 Introduction

The modern enterprise relies heavily on electronicinformation systemsto improve productivity and speed up processes, allowing new service, product development and new business models. As a result, large amount of information is generated, processed, distributed and stored electronically via digital devices and computer networks. However, their vulnerabilities creates opportunities for hostile users to perform malicious activities exposing the underlying critical informationto cyber threats and attacks (Healy at el, 2008; Alharbi at el, 2011).

Currently, finding the most effective way to secure information systems, networks and sensitive data is a challenging task experienced by many organization.The number of potential attackers targeting a given system has increased drastically and the effect of successful attacks have become more serious. For instance loss of fund, lack of confidence from their clients, legal implications and denial of services (Healy at el, 2008; panda Labs 2011). Skilled attackers frequently changes their attacking strategies and devise new methodologiesto negatively affect their existence, amount and quality of evidence generated for analysisin order to defeat the implementedsecurity mechanisms (Garfinkel at el 2007; will at el, 2011). Information Assurance and Security is a crucial component in the corporate environment to ensure that the secrecy of sensitive data is protected, the integrity of important data is not violated, and the availability of critical systems is guaranteed.It plays a key role on nation health, economy and public security and hence continues to be a research area in the pursuit of an efficient, scalable and intelligent system to provide comprehensive security management domain.

Digital forensic is a scientificprocess that facilitates detection of illegal activities and in-appropriate behaviors using scientific tools, techniques and investigation Frameworks which involves diverse digital devices such as computer system, network, mobile and storage devices (Pilli et al., 2010; Rahayu at el 2008). It comprises of a series of steps followed by security experts to obtain accurate and complete evidence which is forensically sound and acceptable in a court of law. The advancement of Internet into a new era and domain such as mobility and Internet of Things, its ever growing user's base and sophisticated cyber-attacks demonstrate the need to deploy advanced IT security infrastructure to handle the current demands in network security (Wang at el, 2010; Maheyzah at el, 2015; Rahayu at el, 2009). Therefore, it is essential to develop a framework that provides tools,techniques and procedures forcollecting, preserving and analyzing large heterogeneous datasets and system's information in a structured way and for supplying detailed and complete information to IT security management in real time.

This work proposes anetwork forensic investigation framework for detecting, predicting and managing cyber-security incidents in a real time multiple sensor environment. The objective will be achieved through a series of steps first by examining existing digital forensic investigation framework. This study allowed us to identify the missing part and the drawback of those systems. The next section will provide the proposed design for an effectiveframework to improve the whole forensic investigation process. Lastly, we conclude the paper and present potential future work

2.0Existing Digital Forensic Investigation Frameworks

Digital forensic approaches are generally categorized into three sections: Integrated Digital Investigation Process (IDIP) Framework, General Network Forensics Approaches and proactive approaches. (Carrier at el, 2003).

2.1 Integrated Digital Investigation Process (IDIP) Framework

IDIP by (Carrier at el, 2003), .is based on the investigation process of a physical crime scene. The framework has seventeen phases which are readiness (operations and infrastructure) phases, deployment (detection and notification and confirmation and authorization) phases, physical crime scene investigation (preservation, survey, documentation, search and collection, reconstruction, and presentation) phases, digital crime scene investigation (preservation, survey, documentation, search and collection, reconstruction, and presentation) phases, and review phase. The main limitations of IDIP based framework depicts the deployment phase which consists of confirmation of the incident as being independent of the physical and digital investigation phase. In practice, this seems impossible to confirm a digital or computer crime unless and until some preliminary physical and digital investigation is carried out. Also it does not offer sufficient specificity and does not draw a clear distinction between investigations at the victim's (secondary crime) scene and those at the scene where the first criminal act occurred (primary source). Neither does it reflect the process of arriving at the latter. Since a computer can be used both as a tool and as a victim .It is common for investigations to be carried out at both ends so that accurate reflections are made. The process of tracing back the suspects seems very challenging when dealing with larger networks.

End-to-End Digital Investigation Process (Carrier at el, 2004), contains nine phases consisting of evidence collection, analysis of individual events, preliminary correlation, event normalizing, event deconfliction (uncountable), second-level correlation, timeline analysis, chain of evidence construction, and corroboration,. It combines the tools of the traditional investigative methods. The focus of the model is on the analysis process, particularly correlation, normalization, and deconfliction of events that are reported from different locations. While the model differs from the other models by the interest it gives to analysis, it does not give enough consideration to evidence searching and finding which a complex and time consuming process is. This model was an advancement as it permits formal verification unlike the preceding models. Any state changes that occurred during the course of the event were clearly represented without providing technical details of the incident.

Incident response to help organizations investigate cybercrimes in a simple manner was developed by (Mandia at el, 2003). The framework consists of seven components: pre-incident preparation, detection of incidents, initial response, and formulation of response strategy, investigation of the incident, 3dczxreporting, and resolution. The analysis phase is included in the investigation component. The framework has limitation since investigation component begins after collecting data from the same components.

Enhanced Integrated Digital Investigation Process framework by (Baryamureeba at el, 2006), consists of five major phases that include sub-phases: readiness (operation and infrastructure readiness), deployment (detection and notification, physical crime scene investigation, digital crime scene investigation, confirmation, and submission), trace back (digital crime scene investigation and authorization), dynamite (physical crime scene investigation, digital crime scene investigation, reconstruction, and communication), and review phase. The approach of the framework classifies the investigation processes into two phases; trace back and dynamite. These phases separate the investigations conducted at the primary and physical crime scenes and depicts the other phases as iterative instead of linear.

Event-based digital forensic investigation framework (Carrier at el, 2003)), is based on the physical crime scene. The framework consists of five phases that include the subphases, i.e., readiness (operation and infrastructure readiness), development (detection and notification and confirmation and authorization), physical crime scene investigation (search and reconstruction), presentation, and digital crime scene investigation phase. Each phase in this framework has a clear goal and requirements to achieve the expected results. The integrated phases, when combined, are insufficient to investigate real cybercrime cases because these phases have not mention the completeness of each phases (Rahayuat el, 2008).

Computer Forensic Field Triage Process framework, (Yong-Dal at el, 2008). It has six phases which include planning, triage, usage or user profiles, chronology or timeline, Internet activity, and case-specific evidence phases. The framework provides the identification, analysis, and interpretation of cybercrime evidence within a short time frame without the need to generate a complete forensic image of the lab. The main limitation experienced by the model is suitability for investigating all types of cybercrimes because evidence is very difficult to distinguish and collect.

Extended model of cybercrime investigation, (Ciardhuáin at el, 2003). Consists of thirteen phases that includes awareness, authorization, planning, notification, search and identification of evidence, collection, transport, storage, examination, hypotheses, presentation, proof or defense, and dissemination activity. This model is more comprehensive than the other IDIP framework because it encompasses almost all the investigation activities but the model needs more evaluation in terms of scalability to ensure that it analyzes evidence efficiently. The model also is based on single-tier processes, focuses on the abstract layer in each phase. The advantage of single-tier processes is that they produce unambiguous outputs. The main limitation of single-tier processes is that they reduce the scalability and flexibility of the investigation when more details are required from the user (Wei at el, 2005).

Hierarchical Framework for Digital Investigations (Beebe at el, 2005), is a multi-tier, hierarchical framework to guide digital investigations. The framework has six phases, namely, preparation, incident response, data collection, data analysis, presentation, and incident closure. The framework introduces objective-based phases and

subphases to each layer in the first tier with the ability to add more details in advance to guide digital investigations, especially in data analysis. The main limitation of this framework is that it is incomplete and requires a more methodical approach to identify the objectives of each layer.

2.2 General Network Forensics Approaches

Evidence Graphs for Network Forensics Analysis (Wei at el, 2010), is a technique for network forensics analysis mechanism that includes effective evidence presentation, manipulation and automated reasoning. The model includes an evidence graph which facilitates the presentation and manipulation of intrusion evidence. For automated evidence analysis, the model has a hierarchical reasoning framework that includes local reasoning and global reasoning. Local reasoning aims to infer the roles of suspicious hosts from local observations. Global Reasoning aims to identify group of strongly correlated hosts in the attack and derive their relationships. The analysis step is the most comprehensive and sophisticated step. There is a need to refine the model in local and global reasoning process with more realistic experiments and also investigate methods to automate the process for hypothesizing missing evidence and validating hypotheses as mentioned by the authors.

Step-by-step framework (Kohn at el, 2006)), Merges the previous frameworks to compile a reasonably complete framework which groups all the existing processes into three stages, namely, preparation, investigation, and presentation, which are implemented as guidelines in network forensics. The aim of the framework is to establish a clear guideline of what steps should be followed in a forensic process. However, understanding how the framework addresses all phases of network forensics in the main stages is very difficult in clarification.

Forensics Zachman (FORZA) (Stephenson at el, 2003) is a framework that focuses on the legal rules and participants in the organization rather than the technical procedures. The framework solves complex problems by integrating the answers with the questions what (the data attributes), why (the motivation), how (the procedures), who (the people involved), where (the location), and when (the time) questions. The FORZA framework includes eight rules: case leader, system or business owner, legal advisor, security or system architect or auditor, digital forensic specialist, digital forensic investigator or system administrator or operator, digital forensic analyst, and legal prosecutor. The main drawback of this framework is that it is human dependent. It requires more tools to conduct a network forensic analysis and to provide accurate results in the investigation phase.

Two-dimensional evidence reliability amplification process model (Khatir at el, 2008), consists of sixteen subphases and grouped into five main phases, namely, initialized, evidence collection, evidence examination or analysis, presentation, and case termination. The phases of the model are described in detail by identifying the roles of the inspector and manager for each phase. The model aims to provide answers to cybercrime questions, such as what happened, when did it happen, and who perpetrated the action, without considering the cybercrime intention and strategy analysis (why and how questions). A similarity exists between incident response and computer forensics (Freiling at el, 2000). The two present a common process model for both incident response and computer forensics to improve the investigation phase. The model includes a set of steps grouped into three main phases, consisting of pre- analysis (detection of incidents, initial response, and formulation of response strategy), analysis (live response, forensic duplication, data recovery, harvesting, reduction, and organization), and post-analysis (report and resolution). Incident response is conducted in the model during the actual analysis. The procedures and methods of incident response are unclear in terms of the type of evidence that is utilized to analyze the incident. No standard method of detecting and collecting evidence exists, which produces insignificant evidence and affects the accuracy of the incident response.

Digital forensics investigation procedure model (Yong-Dal at el, 2008), consists of ten phases: investigation preparation, classifying cybercrime and deciding investigation priority, investigating damaged (victim) digital crime scene, criminal profiling consultant and analysis, tracking suspects, investigating injurer digital crime scene, summoning suspect, additional investigation, writing criminal profiling, and writing report. The model presented the block diagram without any technical details or methods to manipulate with these phases. This indicates that the main focuswas on the number and the type of the network forensics phases rather than how it works and how they conduct the outcomes.

A categorization of investigation process was done (Rahayu at el, 2008) to group and merge the similar activities or processes in five phases that provide the same output. The phases are: Phase 1 (Preparation), Phase 2 (Collection and Preservation), Phase 3 (Examination and Analysis), and Phase 4 (Presentation and Reporting), and Phase 5 (Disseminating the case). The researcher also proposed a mapping process of digital forensic investigation process model to eliminate the redundancy of the process involved in the model and standardize the terms used in achieving the investigation goal.

2.3Proactive Process framework in Network Forensics

Multi-Component View of Digital Forensics (Grobler at el, 2010), includes three components, consisting of ProDF, ActDF), and ReDF. The ProDF component defines and manages the processes and procedures of the comprehensive digital evidence. ActDF includes four subphases: incident response and confirmation, ActDF investigation, event reconstruction, and ActDF termination. ReDF includes six sub phases, which are incident response and confirmation, physical investigation, digital investigation, incident reconstruction, presentation of findings to the management or authorities, dissemination of the result of the investigation, and incident closure.

A theoretical framework to guide the implementation of proactive digital forensics and to ensure the forensic readiness of the evidence available for the investigation process. The framework helps organizations reduce the cost of the investigation process because it provides manageable components and live analysis. The

components proposed in the high-level view make the implementation and automation of the framework more difficult to create automated tools, as stated. Additionally, the process contains phases, such as service restoration, that lie outside the scope of the investigation Alharbi (2011).

Functional Process Model for Proactive and Reactive Digital Forensics, (Alharbi at el, 2011),has two components. The first one is the proactive digital forensic component, which includes five phases: proactive collection, event triggering function, proactive preservation, proactive analysis, and preliminary report. The second component is a reactive digital forensic component that also has five phases: identification, preservation, collection, analysis, and final report. The proposed proactive component is similar to the active component of the multi-component process such that they share the same reactive component process. The examination and analysis phases are combined in the proposed process under a single phase called analysis. The limitation of this framework, it has not yet fully implemented and may be adapted to implementation requirements and it does not address all techniques used by anti-forensics methods, which could affect the ability of the components to resolve the cybercrime in an efficient manner.

2.4 Generic Process Model for Network Forensics

The generic process model for network forensic analysis (Grobler at el, 2010), divides the phases into two groups. The first group relies on actual time and includes five phases: preparation, detection, incident response, collection, and preservation. The four phases in the second group act as post-investigation phases, which include the examination, analysis, investigation, and presentation phase. The first five phases work proactively because they work during the occurrence of the cybercrime saving time and cost during the investigation process. The first phase prepares the network forensic software and legal environments, such as the IDS firewalls, packet analyzer, and authorization privilege. The second phase detects the nature of the attack by generating a set of alerts through the security tools. The third phase extends from the detection phase; it initializes the incident response based on the type of the attack and organizational policy. The fourth phase, which also extends from the detection phase, collects network traffic through suitable hardware and software programs to guarantee the maximum collection of useful evidence. The fifth phase backs up the original data, preserves the hash of all trace data, and prepares a copy of the data for utilization in the analysis phase and other phases.

The other four phases of this model work after the investigation phase and act as a reactive process begin with the examination phase to integrate the trace data and identify the attack indicators; the indicators are then prepared for the analysis phase. The seventh phase is the analysis phase, which reconstructs the attack indicators by soft computing or through statistical or data mining techniques to classify and correlate the attack patterns. The phase aims to clarify the attack intentions and methodology through the attack patterns and provides feedback on how to improve the security tools. The eighth phase is the investigation phase, which aims to identify the path of the attack and the suitable incident response based on the results of the analysis phase. The final phase presents and documents the results, conclusions, and observations about the cybercrime. All the activities of network forensics are included in this model; the present research adopts the phases of this model as a baseline to show how the analysis phase integrates with the other phases.

In generic framework each phase in the first five phases requires a certain amount of time to accomplish its processes. Each phase works in real time; thus, the phases require the same amount of time and processing cost to accomplish their processes. Given that the other four phases work reactively, it is assume that they require more time and processing cost compared with the first five phases. The reason for this assumption is that reactive phases work after the cybercrime happens; therefore, the required amount of time and cost increases during the \investigation process.

3.0 Discussion and Analysis of Digital Forensic Frameworks

3.1 Summary of existing digital forensics framework

All the discussed techniques have their advantages and disadvantageous as summarized in Table 1 below

Table 1: Summary of existing digital forensics framework

Approach	Type	Limitations
event-based digital forensic investigation framework (Carrier at el 2003)	Reactive	the integrated phases, when combined, are insufficient to investigate real cybercrime cases because these phases have not mention the completeness of each phases
Computer Forensic Field Triage Process framework (Yong-Dal at el 2008)	Reactive	evidence is very difficult to distinguish and collect
Hierarchical Framework for Digital Investigations(Beebe at el,2005)	Reactive	It is incomplete and requires a more methodical approach to identify the objectives of each layer.
Step-by-step framework (Kohn at el 2006)	Reactive	Understanding how the framework addresses all phases of network forensics in the main stages is very difficult need clarification.
Forensics ZachmanDigital forensics Investigation Framework (Stephenson at el, 2008)	Reactive	It requires more tools to conduct a network forensic analysis and to provide accurate results in the investigation phase.

Two-Dimensional Evidence Reliability Amplification Process Diagram (Khatir at el 2008)	Reactive	Does not consider the cybercrime intention and strategy analysis (why and how questions)
Common Process Model for Incident Response and Computer Forensics (Freiling at el 2008)	Reactive	No standard method of detecting and collecting evidence exists, which produces insignificant evidence and affects the accuracy of the incident response.
Digital forensics investigation procedure model [31]	Reactive	The model presented the block diagram without any technical details or methods to manipulate with these phases
Mapping process in digital forensic (Rahayu at el 2008)	Hybrid	They did not implement the model
Generic Process Model for Network Forensics (Ricci at el, 2006)	Hybrid	The output of the examination and analysis phase which doesn't mention the methods and techniques which could be used to conduct the output from this phase.
Multi-Component View of Digital Forensics, (Grobler at el 2010)	Hybrid	The components proposed in the high-level view make the implementation and automation of the framework more difficult to create automated tools(Alharbi at el, 2008)
Functional Process Model for Proactive and Reactive Digital Forensics (Alharbi at el, 2008)	Hybrid)	has limited capabilities because it does not include all the anti-forensic techniques,
Cyber Crime Resolving Approach (Mohammad at el 2013)	Hybrid	The modules of the proposed approach were neither discussed nor implemented

From the existing frameworks discussed in the literature review, it is clearly indicated that the digital forensic investigation is a process consisting of several activities although they may be different in terms used and the order followed but they are all designed to achieve similar objective. Also the proposed frameworks are built on the underlying experience to improve the existing ones.

3.2 Design Consideration in Developing Network Forensic Investigation Frameworks

The challenges in current networkforensic Frameworks includes

- The organization tends to develop its own procedures focusing on the technology aspects such as data acquisition or data analysis and hence a change in the underlying technology forces new procedures to be developed hence investigation should be incorporated with the basic procedures in forensic investigation which are preparation, investigation and presentation (Satpathy at el, 2010; Kohn at el 2006).
- The digital evidence is in a disorganized form and as such it can be very difficult to handle and not all of them is obviously readable by human.
- During collection process, the evidence is related to the aspect on how the evidence is searched, collected, analyzed, presented and documented without tampering the evidence and preserving the chain of evidence.
- During the analysis process, the analysis tools used must be legally accepted, performed by experts or qualified person, and the evidence should not be tampered with or lost.
- The huge amount of collected data from heterogeneous devises needs automated techniques to reduce redundancy, and consequently reduce the analysis time and storage requirement of the evidence (Noor at el 2015; Rahayu at el 2008)
- A proactive approach to help response systems react before the network is compromised, and to have the opportunities to overcome the advantages of attacker by predicting the next attacker action as a proactive step (Noor at el. 2015; Grobler at el, 2010),
- The investigation process should discover complete relationship with optimal performance among known and unknown attacks (Maheyzah at el, 2015).
- The approach of presenting and documenting the evidence should be understandable to non-technical person such as jury and judge for example applications of graph, tree diagrams other than text.

4.0 Proposed Network Forensic Investigation Framework

The proposed theoretical framework can be categorized as proactive and reactive as it predict future attacker actions before damage, and automatically respond to attacks in a timely manner The proposed approach includestwo major modules which are linked together with a proactive depository.

i. Online alert collection and preprocessing
ii. Online and offline alert correlationand optimization

The proposed model processes include evidence collection, evidence identification and classification, analysis and investigation. The final phase presents and documents the results, conclusions, and observations about the

cybercrimethese phases are distributed in two modules and linked with the proactive depository.

4.1 Online alert Collection And Preprocessing

The first module gathers alerts from heterogeneous sources in real time, preprocess by normalization and aggregation of alerts based on given feature such as time, IP source, destination address, etc. the intrusion according to level of evidence accuracy so that forensic professionals will have smaller scope of evidence to investigate and analyze. The result will be stored in the evidence depository. The module includes the preparation, evidence collection, and normalization and aggregation phases. This phase improves the investigation process by accurately identifying similar cybercrime cases for investigation.

4.2 Online and Offline Alert Correlation and Optimization

The second module provides an analysis mechanism that includes effective alert correlation to improve the quality of alerts and integrate them with isolated alerts and also construct all possible attack scenarios. This can be done either online and offline mode. It wills also prioritizing intrusion alerts. Evidence graph will be generated to facilitate the presentation and manipulation of intrusion evidence. Based on the evidence graph an automated reasoning mechanism can be developed with the help of soft computing and advanced analytics for automated evidence analysis. The phase aims to identify the attack group, reconstruct attack strategy predict incoming attacks together with their intentions and provides feedback on how to improve the security system.

Table 2: Summary of Processes in the Proposed Framework

Module	Phase name	Activities / processes
Evidence collection and pre process	Preparation	•Attacker Goal Identification and hypothesis formulation •Network Configuration •Privilege Profile and Trust Setting •Vulnerability and Exploit Permission
	Evidence collection and preservation	• Data aggregation from different data sources • Formatting and standardizing intrusion alerts • Improve the quality of alerts through Filtering redundant and invalid alerts. • dimensional reduction
Online and offline alert correlation and optimization	Analysis and examination	Alert analysis through structural, causal and statistical based correlation techniques • Filtering low-interest and false positive intrusion alerts. • Discovering attack scenario. • Verification and prioritizing intrusion alerts. • Forecasting attacker next action. • Forecasting forthcoming attacks
Evidence presentation and dissemination	Presenting and reporting	Preparing and presenting the information resulting from the analysis phase • Determine the issues relevance of the information, its reliability and who can testify to it • Interpret the statistical from analysis phase • Clarify the evidence, and Document the findings • Summarize and provide explanation of conclusions\ • Presenting the physical and digital evidence to a court or corporate management • Attempt to confirm each piece of evidence and each event in the chain each other, independently, evidence or events • Prove the validity of the hypothesis and defend it against criticism and challenge • Communicate relevance findings to a variety of audiences (management, technical personnel, law enforcement)
	Disseminating and documenting	Ensuring physical and digital property is returned to proper owner • Determine how and what criminal evidence must be removed • Reviewing the investigation to identify areas of improvement • Disseminate the information from the investigation • Close out the investigation and preserve knowledge gained

5.0 Conclusion

Digital forensic is a scientific process that facilitates detection of illegal activities and in-appropriate behaviors using scientific proven tools, techniques and investigation frameworks. Existing practices in digital forensic are not scalable and efficient to handle advanced and modern attacks exploiting emerging services resulting from advancement in Information Communication Technology. This research proposed proactive approach in network forensic investigation process that will address the issue of evidence collection and evidence analysis in a real time multiple sensor environment. It is aimed to discover complete relationship with optimal performance among known and unseen/new alerts generated by multiple network sensors in order to improve the quality of alert and recognize attack strategy.

For future work, a prototype will be developed in order to prove the effectiveness of the proposed framework.Various issues will be addressed in the implementation of the new process: the ability to collect and preserve alerts online, predict an attack strategy, optimizing the proactive component throughfiltering false negatives and prioritizing intrusions and predict attack next cause of action andprovide feedbackproactively

.

REFERENCES

1. Healy, L. (2008) Increasing the Likelihood of admissible electronic evidence: Digital log
2. Handling excellence and a forensically aware corporate culture
3. PandaLabs, Annual Report Panda Security's Anti Malware Laboratory 2009. 2010, Panda Security
4. Will, J. P. (2011). 7 - Cyber X: Criminal Syndicates, Nation States, Subnational Entities, and Beyond. In *Cybercrime and Espionage* (Vol. ISBN 9781597496131, pp. 115-133). Syngress, Boston.
5. S. Garfinkel, "Anti-forensics: Techniques, detection and countermeasures," in 2nd International Conference on i-Warfare and Security, 2007, p. 77.
6. Alharbi, S. e. (2011). The Proactive and Reactive Digital Forensics Investigation Process International Journal of Security and Its Applications Vol. 5 No. 4, October, 2011
7. Orebaugh, "Proactive forensics," Journal of Digital Forensic Practice, vol. 1, p. 37, 2006.
8. Palmer, G. (2001, a). A Road Map for Digital Forensic Research. Utica, New York,: Report From the First Digital Forensic Research Workshop (DFRWS).
9. Palmer. G. (2001, b). *A Road Map for Digital Forensic Research.* Utica, New York: DFRWS TECHNICAL REPORT.
10. Mukkamala, S. S. (2003). Identifying significant features fornetwork forensic analysis using artificial intelligent techniques. *International Journal of Digital Evidence*, 1-10.
11. Nikkel. (2005). Generalizing sources of live network evidence. *Digital Investigation (The International Journal of Digital Forensics & Incident Response*, 193–200.
12. Ren. (2004). on a network forensics model for information security. *In Proceedings of the third international conference on information systems technology and its applications (ISTA 2004)*, 29–34.
13. Wei, R. a. (2005). Modeling the network forensics behaviors. In Security and Privacy for Emerging Areas in Communication Networks.*Workshop of the 1st International Conference on. 2005*.
14. SitiRahayuSelamat, R. S. (2008). Mapping Process of Digital Forensic Investigation Framework. *IJCSNS International Journal of Computer Science and Network Security*, Vol. 8(No. 10): p. 163-169.
15. Will, J. P. (2011). 7 - Cyber X: Criminal Syndicates, Nation States, Subnational Entities, and Beyond. In *Cybercrime and Espionage* (Vol. ISBN 9781597496131, pp. 115-133). Syngress, Boston.
16. Siebert, E. (2010). *The Case for Security Information and Event Management (SIEM) in Proactive Network Defense.*SolarWinds.
17. Khurana H, B. J. (2009). A framework for collaborative incident response and investigation. *In: Proceedings of the eighth symposium on identity and trust on the Internet,Maryland;* , 38–51.
18. Reith M, C. G. (2002.). An Examination of Digital Forensic Models. *International Journal of Digital Evidence*, 1(3): p. 12.
19. Mohammad Rasmi, AmanJantan, Hani Al-Mimi A New Approach For Resolving Cyber Crime In Network Forensics Based On Generic Process Model ICIT 2013 The 6th International Conference on Information Technology
20. Baryamureeba and F. Tushabe, "The Enhanced Digital Investigation Process Model," sAsian Journal of Information Technology, vol. 5, pp. 790-794, 2006.
21. Beebe, N. a. (2005). A hierarchical, objectives-based framework for the digital investigations process. *International Journal of Digital Investigation*, 2(2): p. 147-167.
22. Carrier, B. a. (2003). Getting Physical with the Digital Investigation Process. *International Journal of Digital Evidence*, 2(2): p. 20.
23. Ciardhuáin, S. (2003). An Extended Model of Cybercrime Investigations. *fxInternational Journal of Digital Evidence* , 3(1): p. 1-22.
24. Freiling, F. a. (2007). A Common Process Model for Incident Response and Computer Forensics. *IT Incident Management and IT Forensics.* Germany.
25. Grobler, C. C. (2010). A Multi-component View of Digital Forensics. In Availability, Reliability, and Security,.*ARES '10 International Conference.* ARES.
26. Khatir, M. S. (2008). Two-Dimensional Evidence Reliability Amplification Process Model for Digital Forensics. In Digital Forensics and

Incident Analysis.*WDFIA '08. Third International Annual Workshop on. 2008.*

27. Kohn, M. J. (2006). Framework for a digital forensic investigation,.*Information Security South Africa (ISSA).* South Africa: Insight to Foresight.
28. Louwrens, C., & von Solms, S. (2010). A Multi-component View of Digital Forensics. *IEEE Xplore* , 647 - 652 .
29. Mandia, K. a. (2003). Incident response and computer forensics. *McGraw-Hill/Osborne.* , 507.
30. Pilli, E. R. (2010). Network forensic frameworks: Survey and research challenges. *Journal of Elsevier Ltd. 2010.*
31. Ricci S.C, I. F. (2006). Digital forensics investigation framework that incorporate legal issues. *Digital Investigation*, 29-36.
32. Rogers, M. e. (2006). Computer Forensics Field Triage Process Model. *Journal of Digital Forensics, Security and Law,* Vol. 1(2), 9-37.
33. Stephenson, P. (2003). *A Comprehensive Approach To Digital Incident* .

Investigation,Information Security Technical Report,E.A. Technology, Editor p. 42-54.

34. Yong-Dal, S. (2008). New Digital Forensics Investigation Procedure Model. In Networked Computing and Advanced Information Management,.*NCM '08.*
35. Noora Al Khater , Richard E Overill (2015), Forensic Network Traffic Analysis, Proceedings of The Second International Conference on Digital Security and Forensics, Cape Town, South Africa
36. S. Satpathy, S. K. Pradhan and B. B. Ray, 2010 "A Digital Investigation Tool based on Data Fusion in Management of Cyber Security Systems," International Journal of In- formation Technology and Knowledge Management
37. MaheyzahMdSiraj, Hashim Hussein TahaAlbasheer and Mazura Mat Din,2015,Towards Predictive Real-time Multi-sensors Intrusion Alert Correlation Framework Indian Journal of Science and Technology, Vol 8(12)

International Journal of Computer Applications Technology and Research
Volume 5– Issue 5, 312 - 319, 2016, ISSN:- 2319–8656

Presenting an Algorithm for Tasks Scheduling in Grid Environment along with Increasing Efficiency by using Fuzzy Models

Abolfazl Jafari
Department Of Computer Engineering, Islamic Azad University. Sari, Iran

Homayoun Motameni
Department Of Computer Engineering, Islamic Azad University. Sari, Iran

Alireza Ghonoodi
Department Of Computer Engineering, Islamic Azad University. Sari, Iran

Abstract: Nowadays, human faces with huge data. With regard to expansion of computer technology and detectors, some terabytes are produced. In order to response to this demand, grid computing is considered as one of the most important research fields. Grid technology and concepts were used to provide resource subscription between scientific units. The purpose was using resources of grid environment to solve complex problems.
In this paper, a new algorithm based on Mamdani fuzzy system has been proposed for tasks scheduling in computing grid. Mamdani fuzzy algorithm is a new technique measuring criteria by using membership functions. In this paper, our considered criterion is response time. The results of proposed algorithm implemented on grid systems indicate priority of the proposed method in terms of validation criteria of scheduling algorithms like ending time of the task and etc. Also, efficiency increases considerably.
Keywords: Grid scheduling, Mamdani fuzzy system, execution time, membership functions

1. INTRODUCTION

The computing grid is a software and hardware ultra structure and it provides stable, universal, cheap, compatible and reliable accessing to computing resources of the network. In grid environment, tasks are not individually executed in a system. These tasks are divided into subtasks, and each of them is transferred to resources that are the member of grid for execution, Resources of the network are connected to each other by using connection links. A link provides communication and information exchange between two related computers. RMS communicates with all network computers in star topology. The task of RMS is to distribute subtasks between computers and to receive their results. Suppose that we have m machines M_j=(j=1,…,), and these machines process n tasks in the form of J_i=(j=1,…,n). In this way, a scheduler for j task is dedication of one or more machines in various times to execute that task.

There are two sets in grid environment scheduling. as follows:

1) Tasks set, 2) computing elements. The purpose of scheduling algorithm is to schedule applications on computing resources transferred by various users to grid systems.

Tasks scheduling in grid involves three basic stages:

The first stage: resources detection that is accessible resource detection. In this step, resources used to solve the problems are detected and identified.

The second stage: Information collection. In this step, information like speed of resource computing, its failure rate, communication like band width of resource, resources management system and etc are collected by schedules. Then, the best set involving resources and tasks; are selected according to tasks characteristics and related information.

The third stage: Task execution. This stage involves execution of task, collecting the results and finally cleaning up the memory and being prepared for next scheduling. The policies of tasks scheduling are divided into two main groups. The first group is system-oriented group, while purpose is to increase throughout of the system to decrease response time of whole system and to create a load balance or combining them. These kinds of scheduling algorithms have high performance Unpredictable, insecure and dynamic environment shares different services between various users. Due to heterogeneous and dynamic nature of grid, methods used in old systems cannot be used for grid scheduling, so new methods must be taken into account. The research and studies show that revelation optimization methods inspired from the nature have better

effect and efficiency than other methods. Bat algorithm is a evolutionary model based on algorithms inspired from the nature. This algorithm is used to solve optimization problems. The proposed methods dynamically provide an optimized schedule to complete transferred tasks with minimum time of flowtime and makes pan.

2. LITERATURE REVIEW

kang presented grid scheduling algorithm on the basis of self-adjustable tabu search in 2010. The proposed algorithm is suitable to reduce makespan of transferred tasks. In the proposed method, during scheduling time, the panel signs and keeps the list of local optimized solutions in future search process to obtain a simple search method for these solutions. In the proposed algorithms resources utilization is not considered the most important function in tabu search algorithm is neighboring function. In 2012, Egrawal proposed scheduling algorithm to minimize makes pan. The presented algorithm is based on standard genetic algorithm. This method requires a coding design, and should show all possible solutions for scheduling problem. Each special solution can be shown by a chromosome (scheduler) chromosomes are manipulated by two genetic operators and different methods until it stops. In order to manipulate it appropriately, a fitness function is required. In the proposed method it is supplied that tasks are independent.

In 2009, Chang proposed scheduling algorithm to minimize MAKESPAN and loud balance. The proposed algorithm has all characteristics of ants optimization algorithm, and it is considered to decrease ending time of tasks by considering the load of each resource in ionic grid environment, and it is used in Taiwan University. This algorithm changes pheromone intensity according to resources conditions by applying the function of updating local and global pheromone, and it tries to minimize ending time, while the system loud balance is kept.

Izakian suggested optimization algorithm of discrete particles swarm to solve the problem of tasks scheduling in 2010. In the proposed algorithm, minimization of Flowtime and Makespan is simultaneously taken into account. In this method, optimization algorithm of binary particles swarm has been used. The matrix element, are considered as zero and one. Each particle can be converted from Zero to one and vise versa.

In 2009, Chen presented PSO-SA involving combination of optimization algorithm of particles swarm and gradual simulation to reduce MAKESPAN time in grid tasks scheduling. One of the main problems of optimization algorithm of particles swarm is to be deceived by local optimization. In order to solve this problems, gradual simulation that is considered as local search algorithms.

Krouz presented GA-SA algorithm that is a combination of gradual simulation and genetic algorithm in 2010 to decrease MAKESPAN time in grid tasks scheduling. Since genetic algorithm searches whole problem space, and it performs in weak local search, it has been tried to solve this problem by combining it with thermal simulation that is considered as local algorithms

3. PROPOSED ALGORITHM

By using the presented algorithm, Tasks of grid network can be appropriately scheduled. The difference between the presented method and previous methods is that both response time and cost can be considered in this method. In other words, by providing a balance between response time and cost of task, scheduling can be performed in the best way. The results show that, in this proposed algorithm, criteria for measuring the quality of scheduling algorithm (completion time, waiting time and etc) can be improved in comparison to other precisions methods.

3.1 Proposed algorithm description

Since some parameters may be described indefinitely to schedule tasks, they should be designed in mechanism may so that these variables can be analyzed in tasks scheduling. One of variables can be analyzed in tasks scheduling. One of the methods to create this mechanism, Mamdani fuzzy system has been used and considered in this research. In our proposed algorithm, Mamdani system uses membership functions to convert input values to Fuzzy equivalents. Membership function is a function by which input data are converted to fuzzy data; that is, each input in fuzzy system is converted to the number from one to Zero. Membership function is defined for both input and output data. There are various types of membership functions such as trapezoidal, triangular, sigmoid and Gaussian membership functions. In this research, trapezoidal membership function is used to convert indefinite input parameters. Since input variables of the system should be converted to fuzzy values, system input variables should be firstly explained as follows:

3.2 Input variables

1- In order to obtain the response time in a system, criteria of processes priority, computing power of processes and input and output operation speed are used. Now we explain these criteria.

Processing priority: This criterion identifies priority of each processing for execution. In order to measure processing priority two criteria involving burst time and real priority of processing are used. After inserting above values, these values are converted to Fuzzy equivalents by using membership functions.

$$\mu_b = 1 - \frac{\textit{actual busrt time}}{\textbf{maximum } \textit{burst time}+1}$$

$$\mu_p = \frac{\textit{actual periority}}{\textbf{maximum } \textit{periority}+1}$$

After measuring above mentioned values, new priority is introduced as follows. In other words, the process whose processing time requires less execution time and the process that has higher real priority are executed soon.

$$\mu_{new} = \max(\mu_b, \mu_p)$$

After detecting and identifying the new priority for each process, by using "linguistic Variables" of low, intermediate and high: this criterion is converted to equivalent linguistic variable. Diagram of membership function of priority parameter is as follows.

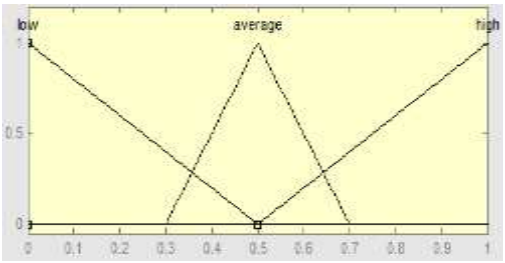


Figure. 1 Diagram of membership function of priority parameter

2- Next input criterion of Mamdani inference system of response time is processors' power. It's clear that if computing power of processors is high, they will be executed quicker, and finally the response time is low, In order to state this parameter by using linguistic variables, variables of 'weak, normal and strong' are used. Diagram of membership function of processors'' power is as follows.

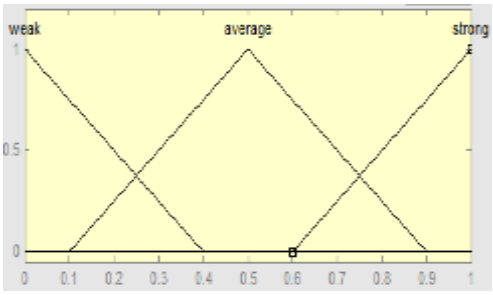


Figure. 2. Diagram of membership function of processors; power

3) Next input criterion is execution speed of input and output operations. They involve operations of entering a process to the system (loading in the memory, and time between service completion and exiting it from the system. If these operations are performed with higher speed, response time is less. In order to state this parameter, linguistic variables including "low, normal and quick" are used. Following feature shows membership function of input and output operations speed parameter.

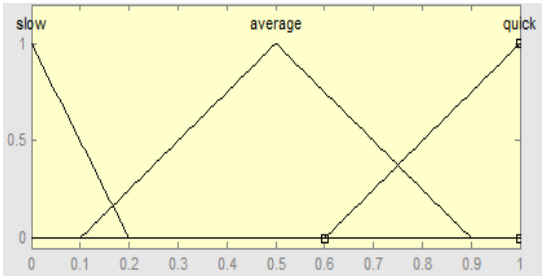


Figure. 3. Diagram of membership function of execution speed in input and output operations

4) After identifying input membership functions, output membership function must be determined as well. In this way, the result of each rule is stated. Output of this algorithm belongs to five intervals between Zero and one, and they include "low, relatively low, intermediate, relatively high, high". The following feature shows output fuzzy membership function.

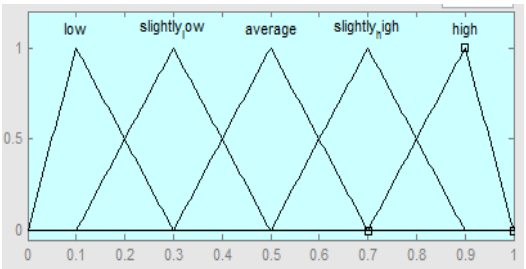


Figure. 4 . Diagram of membership function Response Time

In the second stage, rules database is created. Generally, it can be said that, in fuzzy inference stage, linguistic variables are executed on the basis of input linguistic variables, system rules and membership function, (fuzzy values). In Mamdani method, the methods of "multiplication-maximum" and "minimum-maximum" are used. In our proposed method, "maximum-minimum" method is used for fuzzy inference. The formula used for this purpose is as follows.

$$\boldsymbol{OR = \mu_{OR}AB(x)} \\ \boldsymbol{= \max[\mu A(x), \mu B(x)]}$$

$$\boldsymbol{AND = \mu_{AND}AB(x)} \\ \boldsymbol{= \min[\mu A(x), \mu B(x)]}$$

After identifying formula, we follow fuzzy rules of knowledge base. According to input criteria and linguistic variables, rules of knowledge base are defined as follows.

Table 1. The rules of knowledge base of Mamdani fuzzy system

Rule	Condition	Response Time
R 1 if	(priority is low AND power of processor is weak AND speed of input/output is slow)	High then
R 2	(priority is low AND power of processor is weak AND speed of input/output is average)	High
R 3	(priority is low AND power of processor is weak AND speed of input/output is high)	Slightly high
R 4	(priority is low AND power of processor is normal AND speed of input/output is slow)	Slightly high
R 5	(priority is low AND power of processor is normal AND speed of input/output is average)	Slightly high
R 6	(priority is low AND power of processor is normal AND speed of input/output is quick)	Average
R 7	(priority is low AND power of processor is strong AND speed of input/output is slow)	Average
R 8	(priority is low AND power of processor is strong AND speed of input/output is average)	Average
R 9	(priority is low AND power of processor is strong AND speed of input/output is quick)	Average
R 10	(priority is average AND power of processor is weak AND speed of input/output is slow)	Average
R 11	(priority is average AND power of processor is weak AND speed of	Average

	input/output is average)	
R 12	(priority is average AND power of processor is weak AND speed of input/output is high)	S lightly low
R 13	(priority is average AND power of processor is normal AND speed of input/output is slow)	A verage
R 14	(priority is average AND power of processor is normal AND speed of input/output is average)	S lightly low
R 15	(priority is average AND power of processor is normal AND speed of input/output is quick)	S lightly low
R 16	(priority is average AND power of processor is strong AND speed of input/output is slow)	S lightly low
R 17	(priority is average AND power of processor is strong AND speed of input/output is average)	S low
R 18	(priority is average AND power of processor is strong AND speed of	S low

	input/output is quick)	
R 19	(priority is high AND power of processor is weak AND speed of input/output is slow)	A verage
R 20	(priority is high AND power of processor is weak AND speed of input/output is average)	A verage
R 21	(priority is high AND power of processor is weak AND speed of input/output is high)	S lightly low
R 22	(priority is high AND power of processor is normal AND speed of input/output is slow)	A verage
R 23	(priority is high AND power of processor is normal AND speed of input/output is average)	A verage
R 24	(priority is high AND power of processor is normal AND speed of input/output is quick)	S lightly low
R 25	(priority is high AND power of processor is strong AND speed of	S lightly low

	input/output is slow)	
R 26	(priority is high AND power of processor is strong AND speed of input/output is average)	L ow
R 27	(priority is high AND power of processor is strong AND speed of input/output is quick)	L ow

Now, we want to reduce some rules of this knowledge base by using an algorithm; as a result, complexity of proposed Mamdani fuzzy system decreases. Numbers 1-3 are respectively assigned to each of linguistic variables showing criteria. By considering input criteria and linguistic variables of these criteria, matrix is as follows:

Table 2. Converting the rules of knowledge base to numerical equivalent

1	1	1	1
1	1	2	1
1	1	3	2
1	2	1	2
1	2	2	2
1	2	3	3
1	3	1	3
1	3	2	3
1	3	3	3
2	1	1	3
2	1	2	3
2	1	3	4
2	2	1	3
2	2	2	4
2	2	3	4
2	3	1	4
2	3	2	5
2	3	3	5
3	1	1	3
3	1	2	3
3	1	3	4
3	2	1	3
3	2	2	3
3	2	3	4
3	3	1	4
3	3	2	5
3	3	3	5

As it is observer in this table, the first-third column of the left side shows criteria values of Mamdani fuzzy system. The last column shows output criterion values of Mamdani fuzzy system. Now, we want to simplify these rules by using karnaugh map. The following karnaugh map demonstrates knowledge base rules of Mamdani fuzzy systems.

karnaugh map of the proposed Mamdani fuzzy system

1	1	2	2	2	3	3	3	3
3	3	4	3	4	4	4	5	5
3	3	4	3	3	4	4	5	5

In this table, the rows show the first criteria, table the columns indicate the second and third criterion, Also are want to simplify the above mentioned table on the basis of karnaugh map. As an example, if we are going to simplify some cells of the table involving the value of 2, we consider the following mode:

Table 3. Classification of karnaugh map with numerical values of 2

X_1 \ X_2X_3								
1	1	2	2	2	3	3	3	3
3	3	4	3	4	4	4	5	5
3	3	4	3	3	4	4	5	5

If((x_1=1) AND [(x_2=1) OR(x_2=2)] AND [(x_3=1)OR (x_3=2)OR (x_3=3)]) THEN X_{final}=2

In this rule, X_1-X_3 shows input criteria, and X_{final} demonstrated output parameter.

Table 4. The rules of knowledge base after simplification

Rule	Condition	Response Time	Rule
R1	If((x1=1)AND (x2=1)AND(x3=1 OR 2))	HIGH	R1
R2	If((x1=1) AND [(x2=1) OR(x2=2)] AND [(x3=1)OR (x3=2)OR (x3=3)]) THEN	SLOWLY HIGH	R2

According to the rules obtained from above table, it is observed that the rules reduce from 27 to 11 rules, and in this case, complexity of Mamdani system decreases. For example, we consider the action of and as 8 units, then we have:

For 27 rules: Execution time 27*2*2=108

For 11 rules: Execution time=22*2+21*1=65

3.3 Case Study

Now, we want to implement our proposed algorithm on a system, and measure reliability and response time in the mentioned methods. In case studies, we try to find small and comprehensive examples. In this case, there isn't any complexity, and it is a small sample of other big and real examples. The example of bank ATM lacks much complexity, and it involves architecture products. In this example, appropriate architecture frame of CαISR has been taken into account. We compute reliability and response time by using the proposed method and executable model simulation.

3.4 Implementation

The following table shows the values of input criteria in the proposed system for processes.

Table 5. The values of each input parameters ATM system

ET	A.P	P.O.P	S.I/O
18	1	3	**4**
2	3	7	**9**
1	2	2	**2**
4	6	9	**5**
3	5	6	**3**
12	11	8	**7**
13	7	1	**8**

ET is execution time of the process, A.P is real priority of the process, P.O.P refers to the processor's power and S.I/O points to execution speed of input and output operations. Also, it is supposed that all processes enter the system in zero time.

After entering required parameters of proposed Mamdani system, we should determine the priority of processes for execution by using the mentioned membership function of the algorithm. The following table shows this procedure as well as possible.

Table 6. converting each value of input parameters to fuzzy equivalents

μ_b	μ_p	μ_{new}	**Linguistic variable for new priority**
0.0528	0.0833	0.083	**Low**
0.8948	0.25	0.895	**High**
0.9474	0.166	0.95	**High**
0.7895	0.5	0.79	**High**
0.8422	0.4166	0.84	**High**
0.3685	0.9166	0.917	**High**
0.3158	0.5833	0.58	**Average**

After obtaining a new priority for processes, execution procedure of these processes is as follows:

P3,P6,P2,P5,P4,P7,P1

After identifying execution procedure of processes, the response time is as follows.

Waiting time= 14.86

Response time= 33

After determining input parameters positions, we follow the heart of Mamdani system called inference system of knowledge base. After implementing Mamdani system in CPN, we use MATLAB to obtain output results and to display output membership functions by using input membership functions.

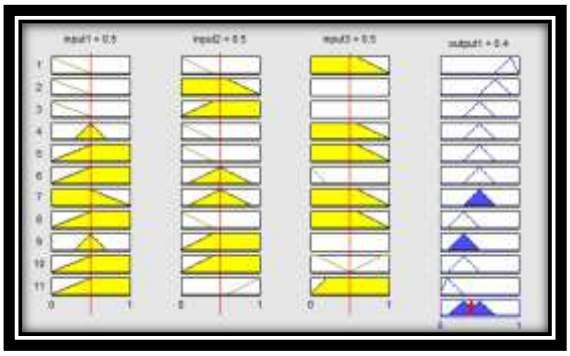


Figure 5. Obtained diagram for response time by implementing the system on MATLAB

As it can be observed in the figure, we can measure response time of ATM system by using simple Mamdani fuzzy system.

4. CONCLUSION

Computing grids provide inexpensive and reliable accessing to others, computing resources. These recourses are heterogeneous, and they are distributed. Also, they are commonly used. On the other side, resources in grid belong to different organizations having managerial policies, models and costs for various users in different time. In this way, owners and users of resources have different objectives, strategies and different supply and demands. In order to manage such a complex environment, common methods trying to optimize efficiency in system level cannot be used to manage the resources.

In this thesis, a new method has been presented to schedule tasks in computing grid environment. In this method applied on the basis of Mamdani fuzzy system, execution time parameter is considered to improve efficiency and performance. In this algorithm, parameters required to evaluate the response time are firstly calculated by using membership functions in Mamdani fuzzy system. Then, the criterion of response time is measured for Mamdani fuzzy system by using output MATLAB software. The results indicate priority of the proposed algorithm in comparison with other previous algorithms.

5. SUGGESTIONS

One of the ideas stated about heuristic algorithm is to study and present new scheduling methods combining other parameters like reliability, load balance and etc and obtaining interesting results. In terms of using Mamdani fuzzy system algorithm, it can be proposed that this algorithm is used in combination with other algorithms. Another suggestion is using some mechanisms for applications classification and dividing the problems to subtasks. If we can do this task by a mechanism with high precision and speed and by considering tasks duration as a determinant of load balance in distributed systems especially grid system, then scheduling precision and speed will be high. In this case, resources are appropriately dedicated, and efficiency and performance increases in grid system.

Future suggestions a terms of scheduling in grid environment are as follows:

1) Studying the methods of error tolerance in proposed algorithms

2) Presenting scheduling in hierarchical order

3) Measuring several parameters in Mamdani fuzzy system

6. REFERENCES

[1] Afzal, A., McGough, A.S., Darlington, J., "Capacity planning and scheduling in Grid computing environment", Journal of Future Generation Computer Systems 24 , pp 404-414, 2008.

[2] BenDaly Hlaoui Y., Jemni BenAyed L.; "Toward anUML-based Composition of Grid Services Workflows",Research Unit of Technologies of Information and Communication, Tunisia, ACM,AUPC'08, July 2008.

[3] . Dai, Y.S., Levitin, G., "Optimal Resource Allocation for Maximizing Performance and Reliability in Tree-Structured Grid Services", IEEE Transaction on Reliability, Vol. 56, No.3, September 2007.

[4] Dai, Y.S., Xie, M., Poh, K.l., "Reliability of grid service systems", Computers & Industrial Engineering 50, pp.130-147, Elsevier, 2006.

[5] Foster, I., Kesselman, C., The Grid 2: Blueprint for a New Computing Infrastructure, Los Alios, Morgan-Kuffman,2003.

[6] G. Murugesan1, An Economic-based Resource Management and Scheduling for Grid Computing Applications, IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 2, No 5, March 2010

[7] Kordic V., Petri Net Theory and Applications, Chapter:Model Checking of Time Petri Nets, Chapter Author:Boucheneb H., Hadjidj R., I-Tech Education and Publishing, Vienna, Austria, First Edition, February 2008.

[8] Levitin, G., Dai, Y.S., "Service reliability and performance in grid system with star topology", Reliability Engineering and System Safety 92, pp. 40-46, Elsevier, 2007.

[9] Li, M., Baker, M., The Grid Core Technologies, John Wiley & Sons Publishing, 2005

[10] Yagoubi, B., Slinani, Y., "Task Load Balancing Strategy for Grid Computing", Journal of Computer Science 3 (3),pp. 186-194, 2007

[11] Saeed Parsa. Fereshteh- Azadi Parand. 2012.Estimation of service reliability and performance in gr id environment: Journal of King Saud Univer sity Engineer ing Sciences vol: 24,pp: 151 157..

Cihan H. Dagli. 2011. Modified SPEA2 for Probabilistic Reliability Assessment: Procedia Computer Science volume 6 ,pp: 435 44.