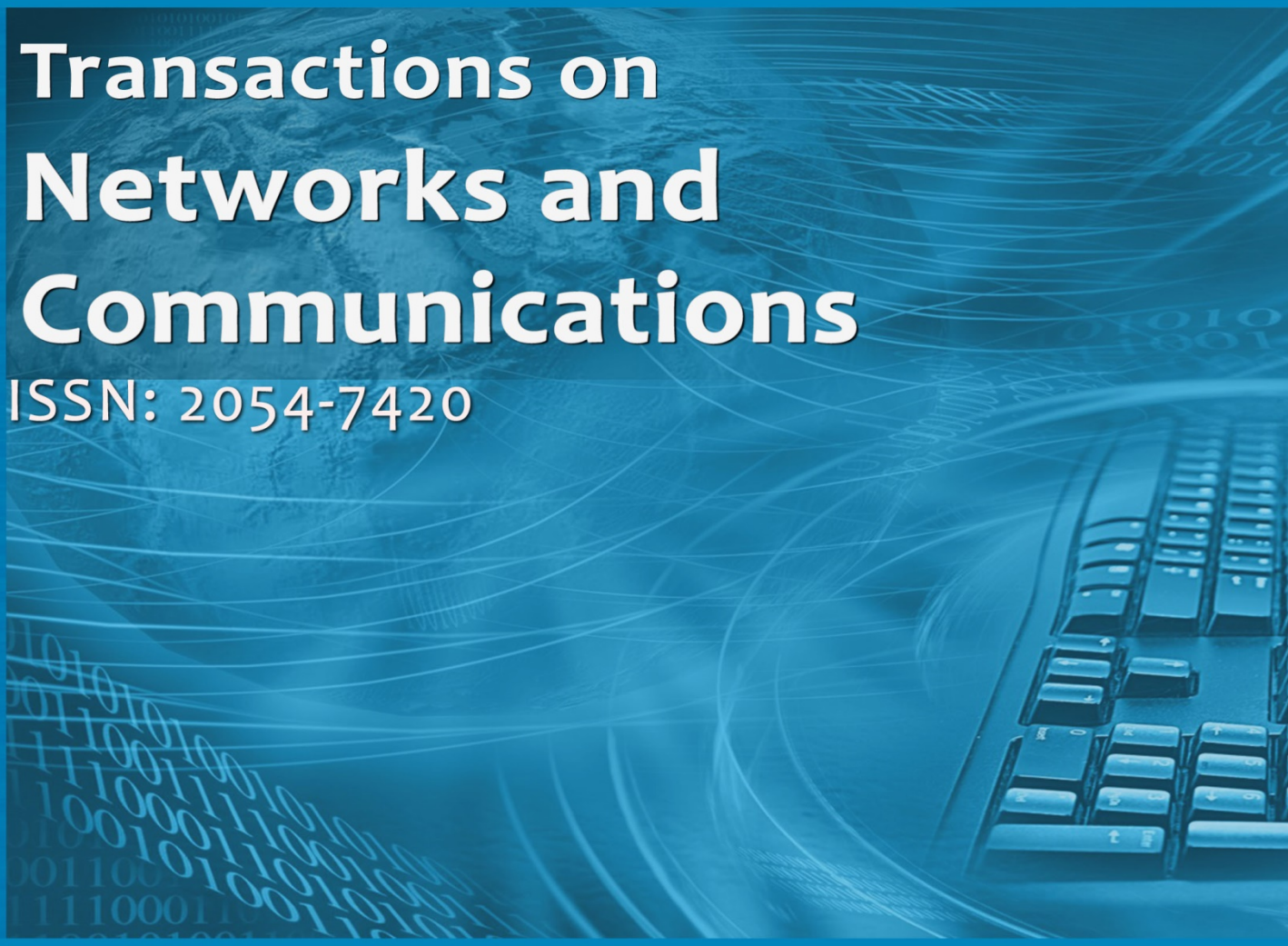
Volume 2. Issue 5 - October 2014

Transactions on Networks and Communications

ISSN: 2054-7420



Society for Sciences and Education - United Kingdom

TABLE OF CONTENTS

EDITORIAL ADVISORY BOARD

DISCLAIMER

All the contributions are published in good faith and intentions to promote and encourage research activities around the globe. The contributions are property of their respective authors/owners and the journal is not responsible for any content that hurts someone's views or feelings etc.

TNC Transactions on Networks and Communications
SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
VOLUME 2, ISSUE 5
ISSN: 2054 - 7420

Security issues in RFID Middleware Systems: Proposed EPC implementation for network layer attacks

Arif Sari
School of Applied Sciences, Department of Management Information Systems, European University of Lefke, Lefke, Cyprus;
asari@ieee.org

ABSTRACT

Recently, Radio Frequency Identification (RFID) technology becomes very popular. Especially low-cost RFID tags are widely used in supply chain management. Due to lack of security considerations in simple RFID technology, performance optimization becomes quite important rather than securing the data transmitted over RFID media. Since security holes shown variety in RFID systems, this paper classifies the attacks that occurs in different layer of RFID models. The security enhanced EPC RFID middleware systems that are widely used in organizations and their vulnerabilities against Network Layer attacks are investigated in this research to clarify the actual impact of network layer attacks in RFID systems. This paper investigates the RFID middleware attacks and impact of possible integration of EPCglobal architecture to mitigate such attacks on RFID systems.

Keywords: RFID Security, RFID attacks, classification, EPCglobal middleware systems, network layer attacks.

1 Introduction

Radio-Frequency Identification (RFID) tags become quite popular since organizations are highly spending in implementing security measures to protect their information assets [1]. The RFID is used to describe a system that transmits the identity of an object or a person wirelessly using radio waves [2]. The simplest types of RFID tags are devices that are quite passive, and have no internal power source. This technology is currently used in security access control systems and can therefore be implemented in enhancing internet security within an organization [3]. This is because RFID has long been as an electronic key to control who has access to office building or areas within an organization. Through the automatic data collection, RFID technology can achieve greater visibility and product velocity across supply chains, more efficient inventory management, easier product tracking and monitoring, reduced product counterfeiting. However providing security such a big network is quite difficult and since the design and implementation of RFID systems are addressed the performance optimization, security issues creates a great challenge for the RFID systems. This paper addresses this problem by analyzing RFID network that uses security enhanced middleware systems and under Network layer attacks. In RFID systems, varieties of attacks are available. In the following sections, different types of attacks are also discussed briefly. As it is also well known that because of the lack of security considerations, the new middleware must be addressed in order to define the security problems and solutions. For that reason, ALE and EPC Global

DOI: 10.14738/tnc.25.431
Publication Date: 11th September 2014
URL: http://dx.doi.org/10.14738/tnc.25.431

Network is also discussed that major concepts of RFID middleware systems in the methodology section where Network-Transport Layer attacks impacts are also discussed.

2 Classification of RFID layers

The RFID systems can be classified into different segments in terms of layers. The Figure 1 below illustrates the RFID communication layers.

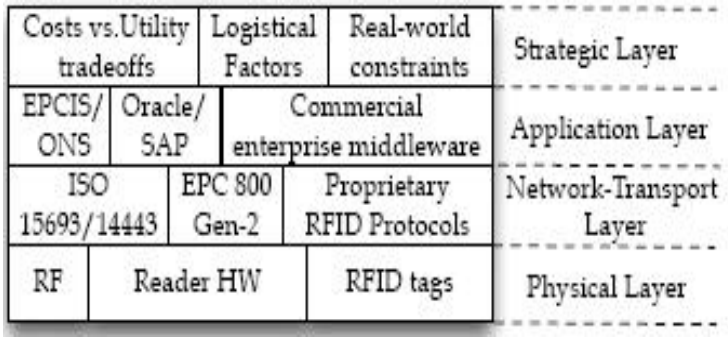


Figure 1: RFID Communication Layers

Due to scope of this study, all of the layers presented on the Figure 1 above are discussed briefly while each and every layer can be investigated separately.

The first layer in the communication protocol is Physical Layer. The physical layer is the combination of physical interface, radio signals and RFID devices. Since the nature wireless communication environment of RFID systems that leads lack of resilience against physical manipulation, the attackers simply disable RFID tags through relay attacks. The Network layer or Transport Layer is the second layer of the RFID communication system that includes all kind of attacks which are related the way that data are transferred between the entities of an RFID network the attacks that are based on the way the RFID systems communicate and the way that data are transferred between the entities of an RFID network components.

The third layer is called Application layer which includes all kind of attacks that target information related to applications and the binging between users and RFID tags.

The Strategic layer includes organizational data coverage area and covers competitive espionage, social engineering, privacy and targeted security threats.

3 Classıfıcatıon of RFID attacks

As it is discussed briefly in the previous section, RFID layers are classified into different layers and each of these layers has its own characteristics. The RFID attacks are classified based on the characteristics of these layers. The Figure 2 below illustrates the classification of RFID attacks.

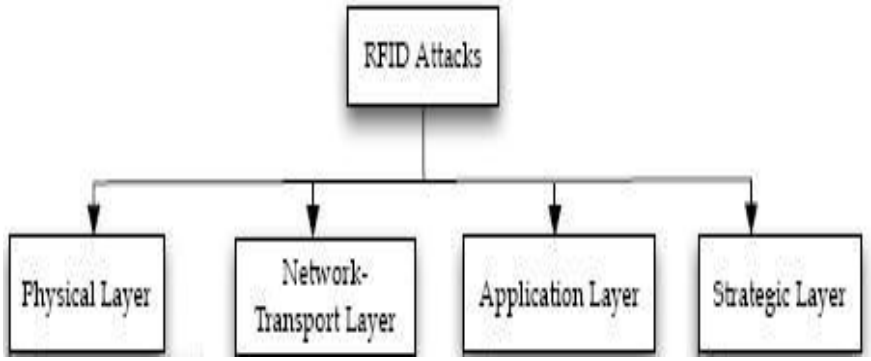


Figure 2: Classification of RFID Attacks

In the literature, varieties of researches have also been stated different classifications of possible threats and risks in RFID networks ([4], [5], [6], [7]).

The detailed classification of business intelligence risks have stated by the author [30]. The researcher [4,7] have proven that the privacy issues of RFID systems cannot be solved through separate studies or separate consideration of layers.

Since this research focuses RFID middleware systems that contains EPC global network, Network-Transport layer must be investigated specifically.

In EPC global architecture, the message that transmitted is secured in the middleware and protected through specific system [8]. This allows companies and organizations to implement and carry out a secure data transmission. In the next section, the EPC global middleware and Network-Transport layer attack is discussed.

4 Proposed Methodology

The RFID Middleware systems contain EPCglobal network architecture. The Figure 3 below illustrates the EPCglobal Network architecture. The EPCglobal Network consists of the ID System (EPC Tags and Readers) EPC Middleware, Electronic Product Code (EPC), Object Name Services (ONS) and EPS Information Services (EPICS) [2]. The EPC sits on the tags and it is a number that is designed to uniquely identify an individual object in the supply chain process. The role of RFID middleware is to handle data interchange between the various systems within the architecture. The diagram below shows an EPCglobal Network Architecture [9].

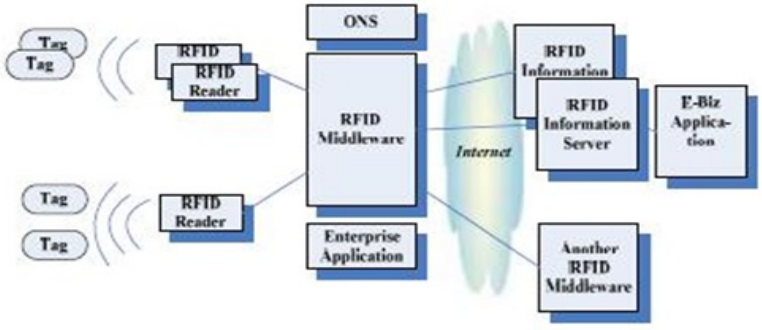


Figure 3: EPCGloblal Network Architecture

All implementation of EPC middleware system belongs to Network-Transport layer, and this layer includes all the attacks that are based on the way the RFID systems communicate and the way that data are transferred between the entities of an RFID network (tags, readers). This section describes the network layer attacks that affect the network transport layer and it's affect on the EPC Global Network architecture and possible solutions to cope with these attacks.

The Secure enhanced middleware system can be achieved through secure middleware system architecture, identification and authentication and transport data protection. This goal has achieved since the proposed mechanism already have the middleware. The classification of attacks shown on the Figure 4 below illustrates that the Network-Transport Layer attacks are categorized under 3 categories which are; Tag attacks, Reader Attacks and Network Protocol Attacks. The Tag attacks are divided into two categories such as Cloning and Spoofing. Cloning is replication of legitimate RFID tags as fake ones that does not require extraordinary financial support and easy to implement through writable and reprogrammable tags. Spoofing is similar to cloning since it's not required to have physical RFID tag, but allows adversary to gain same privileges electronically.

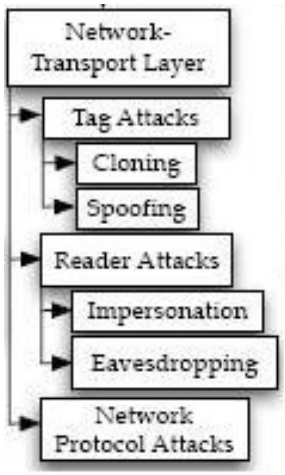


Figure 4: Network Transport Layer Attack Classification

The next category is Reader attacks that fall into 2 different sections such as impersonation and eavesdropping attacks. The Impersonation attacks may occur when the RFID system has unauthenticated communication line and adversary counterfeits the legitimate reader. Eavesdropping attacks sniff the communication between legitimate RFID tags and readers and collect the information. The collected information is used to perform more sophisticated attacks in the future.

The network protocol attacks are the last category that covers the back-end systems and networking infrastructures that communicates with RFID systems. The variety of attacks can be classified and investigated specifically under the Network Protocol attacks since it covers network infrastructure and databases. The operating systems, communication protocols or any other entrusted node especially in wide networks may be used by an adversary to compromise the system.

The cloning and spoofing can be simply prevented in 21st century's technology. A researcher have proposed a specific intrusion detection mechanism against cloning attacks that with the false alarm rates from %2.52 to %8.4 which seems quite successful [10]. The RFID Middleware system encrypts the message through the middleware that prevents passive eavesdropping attacks rather than storing less

information on RFID tags. This system will force users to retrieve requested information from the back-end databases that may lead another information leakage and requires further investigation on Network Protocol attacks. The proposed RFID middleware system uses digital signature encryption functions such as X.509 certificate for authentication or private key [9]. In addition to this, EPCglobal Architecture Application Level Events (ALE) layer uses differentiated access control policies that secure not only the entire transmission media but the message itself at the each intermediary checkpoint. This prevents eavesdropping [3]. In addition to this specific feature, there are other proposed encryption mechanisms proposed by the researchers that can be implemented on RFID middleware systems for to enhance security such as hash-lock [11], randomized hash-lock [12] and chained hashes [13].

5 Conclusion

There are several tasks involved in incorporating RFID in protecting variety of attacks in or outside of the organization. In this study, possible network layer attacks are discussed with EPCglobal network architecture by considering the point of attack based on the RFID layers. However each and every category should to be investigated separately for better understanding of each attack and its countermeasures. The main aim of this empirical investigation was to expose and highlight the network layer attacks on RFID layers. The use of middleware contains encryption mechanisms in its nature so it ensures confidentiality and integrity of the information transmitted over the internet. The study can be expanded by examining also other types of threats and give a better overview of the problem by discussing possible countermeasures in each category of RFID attacks in the future.

REFERENCES

[1]. Kindberg et al. (2002), “People, Places, and Things: Web Presence of the Real World,” ACM Mobile Works & Applications J., pp. 365-376.

[2]. Whiting, R. (2004). “RFID growth poses a data management challenge,” Computing, pp. 29-30. Publisher: VNU Business Publications, UK.

[3]. Finkelzeller, K. (2003). The RFID Handbook, 2nd ed., John Wiley & Sons.

[4]. Garfinkel, S., Juels, A., and Pappu, R. (2005). RFID privacy: An overview of problems and proposed solutions. IEEE Security & Privacy, 3(3), 34-43.

[5]. [5] Ayoade, J. (2007). Privacy and RFID Systems, Roadmap for solving security and privacy concerns in RFID systems. Computer Law & Security Report 23, 555-561.

[6]. Karygiannis, A., Phillips, T., and Tsibertzopoulos, A. (2006). RFID security: A taxonomy of risk. In Proceedings of the 1st International Conference on Communications and Networking in China (ChinaCom'06), October 2006, Beijing, China (pp. 1-7). IEEE Press.

[7]. Avoine, G. & Oechslin, P. (2005). RFID traceability: a multilayer problem. In A.S. Patrick, M. Yung (Eds.), Financial Cryptography and Data Security, 9th International C, FS 2005, Roseau, The Commonwealth of Dominica, Lecture Notes in Computer Science, Security and Cryptology, vol. 3570, (pp.125-140). Berlin, Heidelberg: Springer-Verlag. doi:10.1007/b137875.

[8]. Sari, A. (2010). RFID Security Models use of Security Enhanced RFID Middleware Systems for Enhancing Organizational Data Security. 6th ArchEng International Symposium of European University of Lefke, Vol 6.

[9]. Jieun, S. and Kim, T. (2005). Security Enhanced RFID Middleware System. Retrieved from *http://www.waset.org/journals/waset/v10/v10-16.pdf*

[10]. Mirowski, L. , Hartnett, J. (2007). Deckard: A system to detect change of RFID tag ownership. International Journal of Computer Science and Network Security, 7(7):89 -98.

[11]. Weis, S.A. (2003) Security and privacy in Radio-Frequency Identi_cation devices. Master's thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

[12]. Weis, S., Sarma, S., Rivest, R., and Engels, D. (2003). Security and privacy aspects of low-cost Radio Frequency Identi_cation systems. In D. Hutter, G. M□uller, W. Stephan, M. Ullmann (Eds.), Security in Pervasive Computing, Proceedings of the 1st International Conference in Security in Pervasive Computing, Boppard, Germany, March 12-14, 2003, Lecture Notes in Computer Science, vol. 2802, (pp. 201- 212). Berlin, Heidelberg: Springer Verlag. doi:10.1007/b95124.

[13]. Ohkubo, M., Suzuki, K., and Kinoshita, S. (2003). Cryptographic approach to privacyfriendly" tags. In Proceedings of RFID Privacy Workshop, MIT, MA, USA.

TNC Transactions on Networks and Communications
SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
VOLUME 2, ISSUE 5
ISSN: 2054 - 7420

E-Health Monitoring System using Wireless Sensor Network

Syed Abdur Rauf Magrabi
Global Institute of Engineering and Technology, JNTUH, India
sarmmagrabi@yahoo.com

ABSTRACT

This paper describes about the E-Health monitoring system by using a wireless sensor network. It has wide range of application both in industrial based and as well as commercial based sectors. It mainly deals to monitor the health status of a human body from the harmful sickness and illness like systolic Blood pressure (BP), Heart Rate (HR), Respiratory rate (RR) and temperature of the human body. The necessary equipment and protocols used to monitor the healthy status of the human body are Sensors, communication protocol (TCP/IP), sensor nodes, UML studio and programming in visual basics. The type of platform and the communication based services are used is as follows:

Temperature with Publish-Subscribe based service. The compiler used to adjust the range of values for the patient is VISUAL BASICS. The platform used as a sensor node is CorTemp (Core Body Temperature monitoring system)

Keywords—TCP/IP, Temperature, sensor, UML studio, Visual Basics, Publish-Subscribe based service.

1 Introduction

The general meaning of a Publish is displaying the sender's message rather than addressing the necessary requirements, for the Subscriber it sends the message to all the eligible systems who is asking for the message to receive. In short Publish is message givers (or senders) and Subscribe is message receivers. It is a loosely coupled model in which the senders do not know who their subscribers actually are. The advantages of Publish and subscribe based services are It is a loosely coupled system which does not know who subscribers are, and who are the publishers. Publish and subscribe can continue their work regardless to one another. It is like a client-server based topology. Pub/Sub systems can decouple not only the location of publisher but also the subscribers (and also temporarily). It uses a strategy called a middleware analyst where pub/sub technology can let down the publisher so that subscribers can work; these systems also provide a better scalability than any other based services. The disadvantage of Pub-Sub based services are if one publisher sends any message the subscriber can't detect or make any necessary changes to it. If subscriber does any mistake the publishers will unable to specify or detect that error so more complexity arises when this case occurs [2].

The installation in pub-sub services is quite difficult and both can't handle at the same time, there is no joint for each to them to make any sort of communications in severe conditions.

DOI: 10.14738/tnc.25.382
Publication Date: 11th September 2014
URL: http://dx.doi.org/10.14738/tnc.25.382

2 Block Diagram

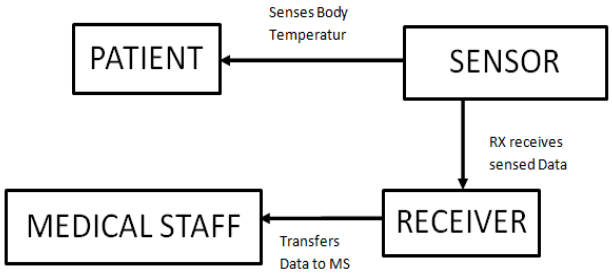


Figure 1: shows the block diagram of E-Health Monitoring System.

This sensor is made up of silicon which contains a micro-battery. When monitoring the system it can be administered into another sensor after the pill has passed. The core temperature sensor is accurate to 0.1 degree centigrade which is also cleared in FDA. Once the pill is swallowed or ingested the crystal sensor vibrates at a frequency relative to the body internal temperature and hence produces a magnetic flux in it and also allows the sensor to transmit a low frequency signal through the patient body. This sensor will be there in human body in between 20-40 hours duration and it will start working in a normal rate without harming the patient body. It can also consist of quartz crystal, communication protocol, insulated coils and circuit board. In wireless sensor network to transmit the data for the core body temperature in a E-Health monitoring system passes through the digestive tract. In this sensor the signal can locate through body of the core temperature and the data recorded by the sensor immediately sends signal to display screen through computer or laptop or any other digital signal. It is specifically designed for human use only 262 kHz. The sensor wirelessly chooses a signal for the conversion of analogue signal to digital signal by using processor technique system. The display of temperature in a real time bases and storage of data for the analysis of the system is done by using monitoring devices. This type of monitoring system is reliable, flexible, easy to use, quick response and data storage [4].

The occupation safety of sensor node mainly concern the hot environment around it with the heat stress is very important issue for workers and it should contributes to higher safety incidents, lower worker productivity, negative economic impact and morale. The frequency for the hot accidents in hot surrounding can cause more moderate environmental conditions. The core body temperature is the most objective measure of heat stress where the system delivers the data to the monitor with the highest degree of accuracy, comfort and ease of use. The absorption of this thermometer pill has a silicon coated exterior with a small battery which is made up of a quartz crystal temperature sensor which can be spaced in a system being used and with the help of circuitry the thermometer pill can react or measure the body temperature in 3 seconds and sends the signal to the system in a short particular of time [4]. This thermometer pill is harmless to human body as like once the pill is swallowed the sensor vibrates at a frequency relative to the body temperature sensor and transmitting low frequency signal through the body within 24 hours of time the pill passes safely from the digestive system without harming the human body.

As shown in figure 1 the necessary steps are useful:

- *The temperature sensor is placed on the body of patient.*
- *The sensor gets waked up and senses the data.*

- *The receiver end receives the data being sensed and transfers the data to Medical staff.*
- *The system is monitored each time and the emergency condition is immediately sent by alarm to Medical staff.*
- *The medical staffs immediate assess the patient.*
- *We use TCP/IP protocol for communication.*
- *The advantages of this health monitoring are the patient is benefited cost wise, light weight and small in size.*
- *This can be practically implemented by Wireless Embedded Technology and Graphical User Interface.*

3 Hardware Used

Features of CorTemp Ingestible Thermometer Pill are described below:

Size: L 0.88" (22.352mm) approximately. Diameter: 0.42" (10.9mm) approximately. Sensor Element is made up of Crystal. Transmission Method: Near field magnetic link Frequency: 262 kHz. Temperature Range: 30 degrees C to 45 degrees C (86 degrees F to 113 degrees F). Accuracy: ± 0.1 degree C. Effective Range: 24" minimum. Power Source: Silver oxide battery. Capsule Material: Dimethyl Polysilicoxane (silicone) Complies w/21CFR177.260 & 175.300 USFDA Regulations. Battery Life: Approx. 7-10 days. Usage: One-time use only: Factory calibrated. Warranty: 90 days [1] and [3].

4 Query Based Information

Query based information contains a specific information injected through a mobile sink the communication between the networks is done in the form of packets. The collection of data through Query based services makes the system more efficient for reducing the problems in a wireless sensor network. The results obtained by these services are in the form of packets. Query information extraction contains limited queries in a system where the computation in the network requires an aggregate range of values in a system. It has spatio-temporal characteristics in the design aspects where the data generation can be possible in a wireless sensor network. The network communication is done by using a TCP/IP based protocol systems which represent higher request for the information based extraction systems. This system should complete the request and response task in a specified duration such that it distributes its information in the query language. The system based information can develop a wide range of access to distribute its values in a sensor node. Where as in the case of publish-subscribe it sends a messages to the system in a messaging pattern and does not functionally programmed to retrieve the signal back to it. In the subscribers messaging it does not program the message and directly sends to the specific receivers directly. The subscribers can express more classes in the receivers messaging systems. This service is like a sibling to each other in a message queue paradigm and it takes a larger part of the message oriented middleware system. This messaging system can elaborate the work done by pub-sub in a message queue system. It is reliable and cost effective method as compare to query based services and also provides a better network scalability with a more enthusiastic networking system. In a query based service the node will be measured by keeping the device on the patient body and publish-subscribe based service the node is measured by putting the device on the patient body. Query based service takes the approach from quality of service based methods with communicate with peripheral devices and hence makes the system simpler than any other services for more information and features refer [6] and [8].

5 Software Design

It stands for Unified Modeling Language. It is an object oriented programing language that can be easy understood by an Engineer/non-Engineer. It involves the software design and analysis of system that deals with use case, class diagram i.e. object interaction, sequence diagram i.e. the flow of data between objects. The people who use this UML are requirement analysis, architect to build such as platform for any kind of work, database professionals, testers and project managers

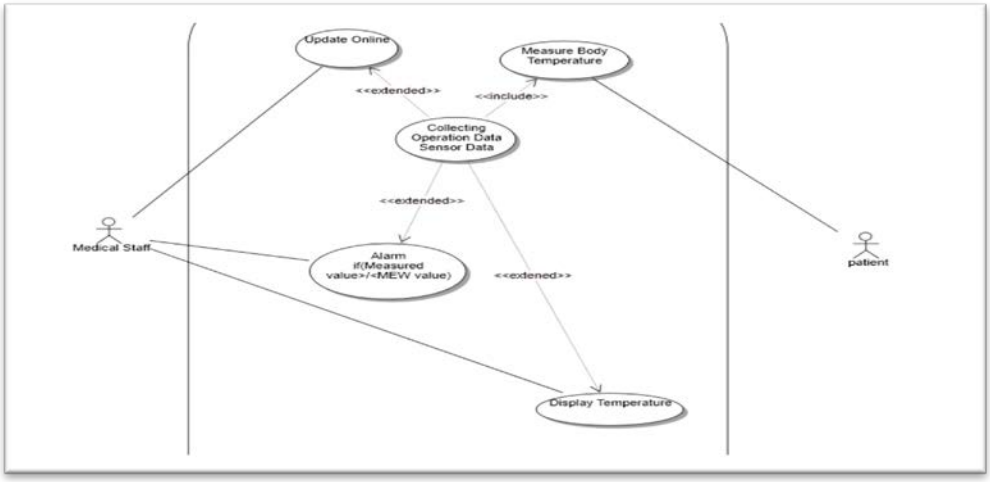


Figure 2: Use Case Diagram.

Use Case gives out the basic idea that how a system can be used. They provide a high level view of a system that can be easily understood by both domain expert and system developer. The use case above shows the interaction between two actors. They are Medical staff and the Patient. The scenarios that take place are as follows:

- The sensor detection: the sensor senses the actual and critical body temperature of the patient and sends the data to the receiver (Medical staff).
- If the sensed data exceeds the threshold value or below the threshold value then there is an alarm sent to the Medical staff stating the emergency of patient.
- Medical staff detects the data received and even the online update can be done by Medical staff.
- The sensed data is displayed on LCD.

The class diagram shows the set of properties and behaviors that are shared by set of objects.
It has three parts the class name, the attribute and its method.
They are shown below.

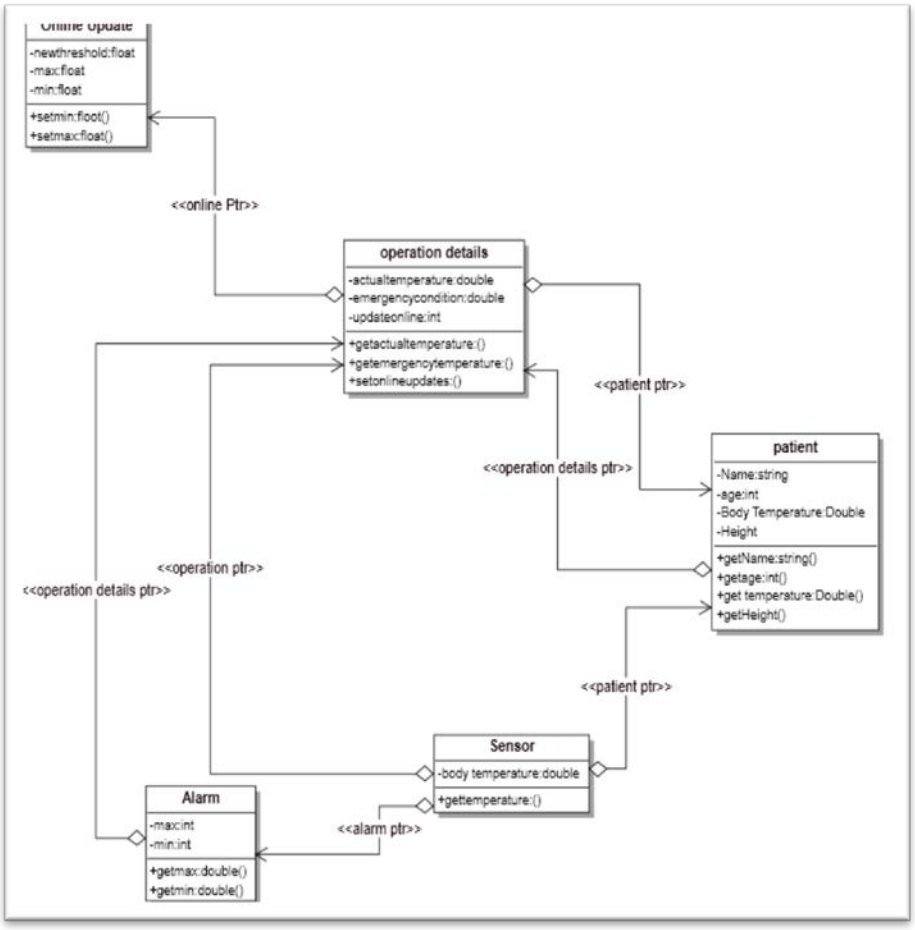


Figure 3: Class Diagram of the system

It is graphical representation of the system basically the scenarios from use case that shows interaction of objects with respect to time.

- Objects that take part are with indicated by vertical line.
- Messages or the data that is passed between two objects are shown in horizontal line.

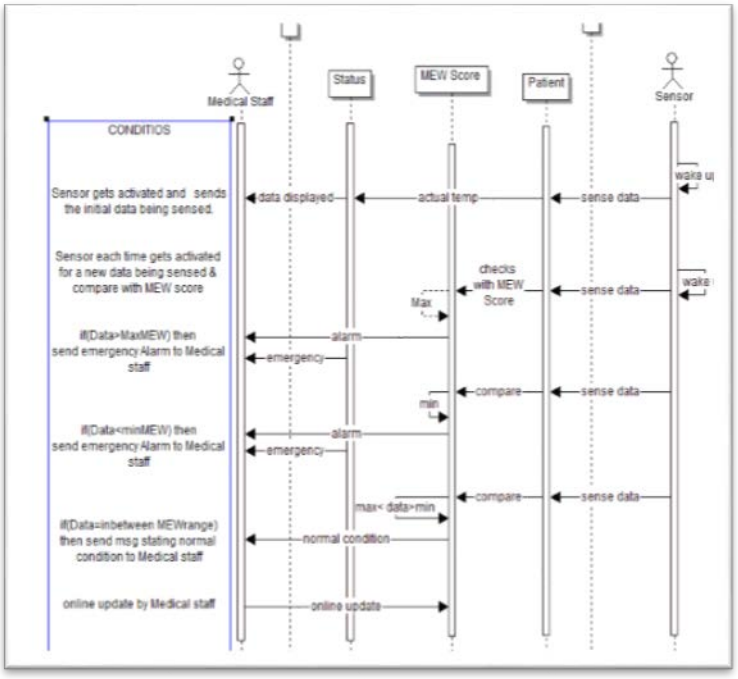


Figure 4: Sequence Diagram

The above sequence explains the action or the condition that takes places between two actors.

6 Conclusion and Results

E-Health Monitoring System is very efficient in design, fast in processing, ease of use for user. This design analysis can also be implemented and extended to monitor the health status of a human body from the harmful sickness and illness like systolic Blood pressure (BP), Heart Rate (HR), and Respiratory rate.

The results for temperature sensing system from an human body is as shown below:

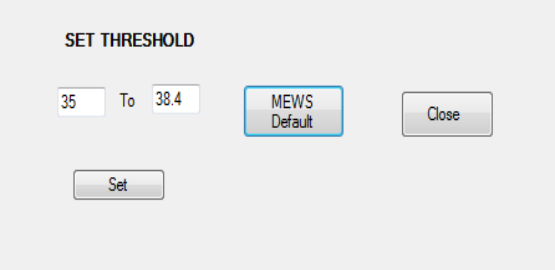


Figure 5: After debugging (Set Threshold before measuring Patient Body Temperature)

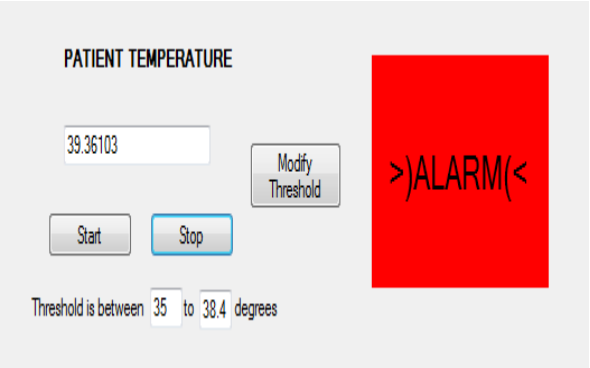


Figure 6: Shows Patient Temperature that exceeds the MEW scores Chat with Alarm

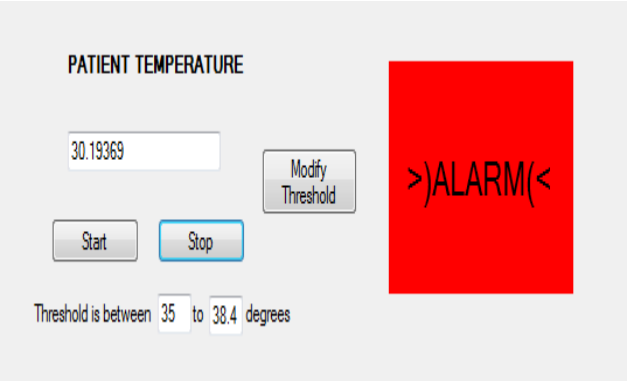


Figure 7: Shows Patient Temperature that is below the MEW scores Chat with Alarm

Figure 7 represents the abnormal temperature of the human body where the threshold value is below than the predetermined value and hence the patient should be needed an intensive care.

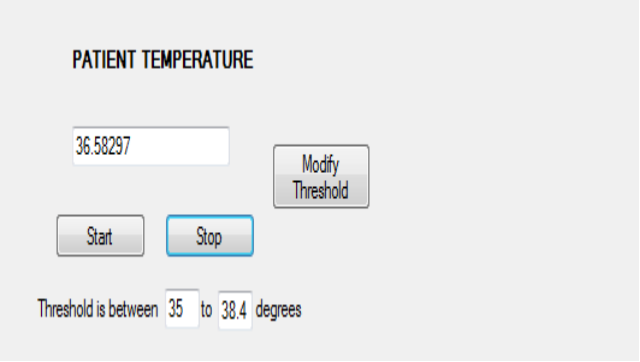


Figure 8: Shows Patient Temperature that is in the range of MEW scores Chat.

Therefore the -Health monitoring system using wireless sensor network is detecting the temperature of the patient body; figure 8 shows the normal temperature of a patient body where the signal detecting the healthy status of human body hence no alarm is generated. The temperature difference is set in a predominant fashion and then modifying the threshold values in the system by using wireless sensor networks. This method of design and hardware implementation can also be done by Python Programming along with hardware equipment known as Onyx II and WristOX_2 [13].

ACKNOSWLEDGMENT

This work was supported and funded by Global Institute of Engineering and Technology (GIET), Hyderabad. Andhra Pradesh. 500075.

REFERENCES

[1]. Rees Scientific Corporation (2010) *Temperature Monitoring*[online] available from: <http://www.reesscientific.com/>[25May2012]

[2]. Class handouts: M06CDE, M01CDE.

[3]. HQInc('n.d.') *Core Temperature Body Monitoring System* [online] available from: <http://www.aesolutions.com.au/images/products_brochure/CorTempBrochure2010-6901.pdf>[26May2012]

[4]. Anon(2009)*Publish Subscribe Model* [online] available from: <http://c2.com/cgi/wiki?PublishSubscribeModel>[28May2012]

[5]. AES('n.d.')*Active Environment Solution* [online] available from: <http://www.aesolutions.com.au/products/heat-stress/core-body-temperature/cortemp.php>[27May2012]

[6]. ACM digital library(2012) Access control in publish subscribe system

[7]. [online] available from<http://dl.acm.org/citation.cfm?id=1385993>[20May2012]

[8]. ACM digital library(2012)Concept based Query expansion

[9]. [online]available from<http://dl.acm.org/citation.cfm?id=160713&bnc=1>[26May2012]

[10]. Procullux Media Ltd(2010) *Publish/Subscribe*[online]available

[11]. from <http://looselycoupled.com/glossary/publish-subscribe>[22May2012]

[12]. Parlez UML(2005)*What is UML*[online]available from<http://www.codemanship.co.uk/parlezuml/e-books/umlformanagers/index.htm>[29May2012]

[13]. *eHealth Remote Patient Monitoring System (nd) available [online] <www.nonin.com/eHealth> [Feb2014]*

TNC Transactions on Networks and Communications
SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
VOLUME 2, ISSUE 5
ISSN: 2054 - 7420

Flexible Pricing Models for Cloud Services

Sururah A. Bello
Obafemi Awolowo University, Ile-Ife, Nigeria
Gazali Abdul Wakil
Computer Science & Engineering Department, Obafemi Awolowo University, Ile-Ife
apinkebello@yahoo.com

ABSTRACT

Cloud Computing's service models IaaS and PaaS have a number of proprietary pricing models as the services has been commoditized to some extent. The SaaS so far has been known with a flat price within the usage time. But pricing models need to be more flexible to prevent customers from thinking that paying same price for a service over a period is no more cost effective, in spite of the level of utilization of the service. For Cloud Computing this has to be taken into account. Africans are not used to fixed price models in their business transactions. In order to expand the acceptability of Cloud Computing to African a means of disputing prices is necessary. This study proposes cloud utility price models to give the Cloud customer the luxury of different usage style and determine a customer specific, individual, most suitable price model. It gives the customer the opportunity to choose a price model for the predicted usage and work within the budget. The proposed cloud utility price models presented using use cases.

Keywords-Cloud Price Models, Utility Price Models, IaaS PaaS, SaaS

1 Introduction

Cloud computing is categorized under utility services because of its various definitions and mode of deployment [1], [2]. Though, Cloud Computing is being described by [3], as a technological change brought about by the convergence of a number of new and existing technologies. The commercial value of use with typically a self-service pay-per-use business model is deeply rooted in Cloud Computing and distinct it remarkably from previous computing paradigms. This created the need for a means to transform the computing services into monetary entities, hence the need for new business models. There is also the need to carefully select an appropriate Business Model as described in [4], as some Business Models have been established to have different success than others. An appropriate Business model is required to translate the technology into service value. Using Business models, IT services are being translated as electricity into General Purpose Technology (GPT). According to [5] Business Models with metered usage models are proposed for products whose standardized quality is to some extent regulated. A determined worth will acquire a certain quantity, weight or other measure of goods and services. The method of setting the worth is pricing. Thus pricing is a very important decision in Cloud Computing services as it requires a number of considerations. Proper pricing required a standard unit of measurement. There is the need to commoditize to a standard level in order to price effectively.

DOI: 10.14738/tnc.25.281
Publication Date: 17th September 2014
URL: http://dx.doi.org/10.14738/tnc.25.281

The three basic cloud computing service models are Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). Typically in IaaS the commodities are CPU power, memory, I/O usage for each virtual machine. For the PaaS in general the provider is commoditizing the availability per month, used storage per month and used bandwidth. PaaS is considered to be enhanced or value added virtual instances and is offered bundled together with the infrastructure. For the SaaS usage time per month is mostly quantified. The existing commoditization techniques do not really portray the true utility model of Cloud systems. In the existing models, for example, once a virtual machine is started, the billing starts irrespective of the actual real usage of a Cloud service. It should be distinguished between a heavily used virtual machine and an idle virtual machine.

This study proposes an enhanced billing system that incorporates other minute entities. For the PaaS there is the possibility of incorporating cost of deployment and customer usage of APIs (testing tools usage). For the SaaS, we intend to extend the monitored entities to the number of users accessing the service, the client transactions, what are clients actually doing on the Cloud service and the number of users per service. More exact would be to charge the customer according the number of users actually using the service not just the length of usage time. The usage history of the cloud customers will also be considered as this will help in retaining existing customers. This study specifically focuses on improving pricing for the PaaS and SaaS.

The rest of the paper is organized as follows. Section 2 is discussing related work. Section 3 analyses price models. Section 4 presented the taxonomy of usage items for monitoring in Cloud services. Section 5 presents the Use Cases using usage graphical interface for controlling the improved price models. Section **Error! Reference source not found.** describes other possible rewarding models that can be employed. Section 7 concludes the paper.

2 Related Work

Business Models in Cloud Computing is an emerging area. Generally speaking, Business Model aligns the components and functions of the business as well as revenue generated and the expenses incurred. In [6] it is stated that Business Models are tools for expressing business logic and describing customer's values. Cloud Business Models can be said to be a tool for capturing economic value from Cloud Computing. Since a Business Model depicts the story that explains how a firm works [7]. Business models must evolve with markets trends, as a model that succeeds today might fail tomorrow. It is difficult to apply the traditional system centric allocation policy in a highly dynamic and distributed environment [8]. Hence the traditional business model cannot be applied in the Cloud System. Transition from the traditional Software Packaged delivery to Software Delivery as a service in Cloud Computing impact a change from the existing Business Model, this was illustrated with a Gaming Software firm in [9].

According to [10] Cloud Computing Business Models are classified into eight types and gave the strength and weakness of each model. It further related the model to a Cloud Cube Model (CCM) and used the Hexagon Model to illustrate the sustainability of an organization through the adoption of the right business model. This paper focuses on the pricing models. The commercial success of Cloud computing strongly rests on a cordial business relationship between the provider and the consumer through a good pricing system. An Auctioning Resource Pricing policy was proposed in [11] where user can predict the

future resource pricing as well as satisfy budget and deadline constraints without knowing offers from other users to solve the price prediction problem. The study favors the cloud provider. In [12] it was established that the prices charged for computation on IaaS is personally and socially unfair. Since the multi-tenancy infrastructure in Cloud Computing allows for VM interference in disk, network I/O and CPU. Users are charged on consumption per hour and consumption has been prolonged by the VM interference hence users are paying more than the actual consumption. The cost is not even the same for different users. Hence proposed a new charging scheme that will make users pay for what is actually being consumed. The study also focuses on better pricing system for the consumer, same direction as ours but specifically for Computation as a Service under the Infrastructure as a Service. The study favors the cloud consumers.

[13] identified four differential pricing tactics that are available to Internet stores: buyer identification, time of purchase, purchase quantity, and asset/usage. [14] discussed pricing mechanisms on the Internet and also agreed that prices can be customized for each buyer according to various rules that involve, for example, customer location, purchase history and click pattern. Hence it may not out of place to apply these strategies to the Cloud Computing market which is also a business on the Internet environment. A friendly pricing scheme will go a long way to retain cloud customers. Generally When consumers perceive price unfairness they feel negative emotions like anger, outrage, disappointment and may not repeat purchase **Error! Reference source not found.**. In the study of pricing in the Cloud in [16] found out that an unfair pricing scheme could foster dissatisfaction with users and could eventually lead to loss of customers. The paper is a deeper description about fine-grained charging units into the SaaS, to improve the pricing system and endear more users to the SaaS. Our study incorporates other minute's entities into the pricing system thereby enticing more users. Thus favoring both the cloud provider and the cloud consumer.

3 Pricing and Cloud Computing

The total utility can be aggregated as the entire satisfaction or benefit of a customer in a given services. Economists employed the consumer demand theory to determine the total utility as the theory studies consumer behavior and satisfaction. Utility is also used to denote services that are required to satisfy fluctuating customer needs. These services are in turn paid for based on usage and not in flat rates. Electrical power and water are basic utilities known for a long time. Later utilities that emerged are public transportation, radio, television, telephone and the Internet access and recently the Computing services under the tag of Cloud Computing. In [17], Cloud is categorically stated as the 5th utility, placing water as the 1st utility, electricity as the 2nd utility, telephone and public transportation in the 3rd category while Radio and Television fell in the 4th category.

According to [18] water, electricity and public transportation employed the metered usage of service business models. While radio, television, telephone and Internet access uses mainly the subscription model with some metered services for special services. Cloud Computing is often seen as a part of larger development towards long-dreamed vision of society where computing is delivered as a utility [19]. [20] imagined the 21st century as where computing is being transformed to commoditized services and delivered as standard utilities such as electricity and telephony.

The cost of compute power can be calculated by amortizing the capital cost over the lifetime of the system as done in costing electricity systems [21]. The total cost of a cloud service will be a function of

capital funds, depreciation, interest rates and system lifetime. In [22] eight categories of cost in a Cloud System are identified, they are: server hardware cost, software cost, support and maintenance cost, network cost, facilities cost, cooling cost, real estate cost. These cost will be translated to the consumer in two basic categories fixed cost and the variable metered cost.

Cloud cost can be broadly divided into fixed cost and variable cost. Fixed cost include, startup cost and availability cost, this is synonymous with the access cost. The variable cost is the operational cost, which is basically based on usage and is determined by the consumption. In [5], the evolving relationships in Cloud usage was observed, and it was discovered that there is a tendency for the customer thinking that paying for a service over time is no more cost effective after a period. Thus there is a need to review the existing price models in Cloud services.

According to [5] the metered usage model is applicable to products whose standardized quality is to some extent regulated. Therefore proposed the metering model for IaaS (processing power, storage devices, servers, I/O, other hardware etc). The IaaS costing to include legal cost (e.g court or litigation cost), data restoration and disaster recovery cost (though a limit can be set) and regulatory requirements (location dependent regulations, regular backups). The metered model of the IaaS can have variants like metering plus subscription or fixed basic plus metered price model to allow for flexibility. Most of the time the infrastructure is bundled with the PaaS, so the PaaS cost is better incorporated as added value in IaaS. In costing PaaS, customization can be allowed like a free start, billing above free limit and metering price model. The subscription and pay as you go model is proposed for the SaaS layer. The subscription can also have considerations for booking in advance and discounts for pre-payment.

Pricing requires that a consideration is given when exchanging or transferring ownerships. Pricing is one of the most critical decisions in introducing new services [23] as the consumer demands, competitors pricing and market trends are to be considered. Pricing mechanisms describes the means by which several decisions are taken by consumers and businesses interact to determine the allocation of scarce resources between competing uses. It is required for optimum pricing. Traditional pricing mechanisms may be inadequate in Cloud Systems due to the dynamicity of the system [24]. Commercial success with Cloud services can only be achieved by developing adequate pricing mechanisms [25]. Next is an overview of possible price models:

1.1 Static:

Pricing mechanisms can be static for a period of time, that is fixed, it's not going to be affected by consumer characteristics be it volume and may not be affected by the market value also. Static pricing has a number of flavors. Customers can be made to pay per use, the payment will be a function of the time or quantity consumed of a specific service.

- *Flat fee:* Customers may also be made to subscribe that is, pay a flat fee for a product or service for a fixed price.
- *Menu Price:* The Menu Price variant expects the customer to pay a price that is already found on the catalog.

1.2 Variable by market value

On the other extreme of the static pricing is a pricing mechanism can be dictated purely by the market value. This pricing mechanism has a number of variants.

- *Haggling or Bargaining:* In Haggling or Bargaining the buyer and the seller dispute the price to finalize on a productive agreement. It allows the customer to capture surplus and allows for price discrimination, as a richer and desperate buyer may pay more.
- *Yield Management:* The Yield Management variant is also a discriminatory pricing. It allows pricing policy for optimizing profit by anticipating, influencing and forecasting customer behavior (airline seats, hotels rooms).
- *Auctioning:* The Auctioning is a variant that offer services or product for competitive and open bid by increasing the price and then selling to the highest bidder i.e. Forward Auction. The Reverse Auction allows many sellers to bid for one buyer s good in decreasing decrements of price.
- *Dynamic Market:* The Dynamic Market flavor results from continual change in both supply and demand of a product or service. No individual buyer or seller (single entity) has the ability to change the price but the buyers and sellers collectively can.

1.3 Variable by customer characteristics

In another category of pricing mechanism, pricing can be determined by the characteristics of the customers, like the volume consumed and other customer preferences but will not be affected by the market value. This is also implemented in a number of ways.

- *Service Bundles:* Prices can be set according to the bundling of services features.
- *Customer CV*: The history of customers using some characteristics can also be used in determining the pricing for a set of customers.
- *Purchased volumes* can as well be used to differentiate prices for different customers.
- The *Customers valuation* can also determine the final price of a product or service.

Business models that are based on variable pricing are known to be thriving in Africa. Africans are not used to static pricing model. In order to expand the acceptability of Cloud Computing to the African market the variable pricing system must be incorporated deeply into Cloud Computing. In order to implement a variable pricing system as against the current static pricing further minutes entities needed to be incorporated.

4 Priced Items in Clouds

Clouds have a pay-per use model, which implies the usage monitoring of Cloud resources in clouds essential. Based on a survey of the monitoring items of Amazon [26], GO Grid [27], RightScale [28], and Salesforce [29], a taxonomy of Cloud Use Items has been developed at the level of IaaS, PaaS, and SaaS. From the provider point of view, everything that causes cost is of interest. From the customer point of view business related values more are important.

4.1 Use Items in IaaS

First an overview of the use items in IaaS, categorized in Compute, Storage, Network resources with their attributes (see Table I). A long way to go is the integration of QoS into the Cloud infrastructure, but has to come, if Clouds want to be successful in business areas.

Table-I: Use Items in IaaS

Compute Resource	
Attributes	**Explanation**
Number of Instances	number of on demand
Type	Depending on the CPU performance (e.g. MIPS), Memory size, etc.
Time of usage	Typically per hour
Operating system	Licensing issue
QoS	Snapshot, backup, reservation, etc.
Storage Resource	
Attributes	**Explanation**
Storage Volume	Volume per month
QoS	Redundancy, backup, etc.
Network Resource	
Attributes	**Explanation**
Message transfer	Upload, download, intercommunication
Components	Need for firewall, routers
QoS	Guarantee of bandwidth, etc.

4.2 Use Items in PaaS

Table II lists the attributes in the service model PaaS, which are not so wide spread as the IaaS ones. Of course all the use items of the IaaS can be taken into account at the PaaS level, but there are additional ones.

Table-II: Use Items in PaaS

Attributes	**Explanation**
Deployment	Data size and time of the deployment
Database	Volume size and performance issues
Scalability	Min/max scalability limits, scalability speed, etc.
Application development	Support for development, debugging, library support, etc.
Programming language	Type of programming language and possible licensing issues
QoS	Availability, support for APIs, security, etc.

4.3 Use Items in SaaS

At present for the SaaS, only the time of usage is being monitored. The focus of this study is on SaaS. For the SaaS, we intend to extend to the monitored entities to

4.4 No of Users Accessing the Service

Monitoring the number of users accessing a service implies service specific inspection. It might be used to adjust the charge for the service. A less popular service will be visible and can be improved upon. A very active service may have it charge slightly raised to increase revenue for the provider and this may also implies decreasing the price for the less active service to entice customer to it.

4.5 No of Real Transactions

This is reflected by measuring users actually hitting the service not just the length of time of usage. Accessing a service may not necessarily means carrying out transactions with it. A customer may log in to the service and be idle, whereas another customer is actually completing a transaction. So price differential can be set in for the two categories of customers. This will be based on the actions carried out with the service. This will also help to focus on key customers that is those that are actually clicking on the service.

4.6 The Usage History

The usage history of a customer could also be used to initiate a price differential. To retain existing customers, long usage history may be rewarded. This may entice a customer to stay with an old client. Economically, it is cheaper to retain existing customers than acquiring new one. An experienced customer requires less help and most likely has fewer problems to deal with. A good retention strategy to reward loyal customer decrease the customer's overhead and also lowers the maintenance cost for the provider. A long term customer will likely purchase more services and may even introduce new customers.

4.7 Period of Usage

The time of usage of some services could be used to determine the price. The energy companies charge differently for energy utilized during the day and during the night. Instead of zero utilization of SaaS services in the night, a good discount for night usage might entice some SME users. Existing customers who are likely to require less help desk attention could greatly benefit from this. The customer will pay less while the provider will have more revenue whereas the service would have generated nothing in the night.

Table III is summarizing the newly introduced monitoring items:

Table: III Use Items in SaaS

Attributes	Explanation
No of users accessing	More users generally cause more load
No of real transactions	Completed transactions
Usage history	Enable rewarding if experience customers
Usage period	Daytime or night usage
QoS	Request Response Time, Web Application Firewall, etc.

5 Use Cases

To illustrate possible applications, this section presents use cases with corresponding sample implementations. Varying prices with respect to users peculiarity will further exhibit the pay as you use

feature of cloud and will endear more users to Cloud computing. Users are presented with the Cloud Price model template. For instance, a cloud user is subcribing to use the Web Server service as shown in Figure 1. The user specifies the number of hours the service is needed and also the period of the day, which is 'day' as shown in the figure. In this case the Web Server is needed for 10 hours during the day period. In figure 2 , the summary of usage and the corresponding price is displayed. The user only used 2 hours out of the 10 hours suscription, there is a discount and no penalty as there was no violation. Once these specifications of usage are made the customers get the price attached to the expected usage.

In case the usage does not correspond with the initial specifications then, penalty will be charged. To demonstrate this, figure 3 shows the subscription details for Database server by the same user while the usage details, taking into account the violation, is shown in figure 4. This study is proposing a flexible utility model where the consumer has the opportunity to adjust the offering giving rise to a number of combination. The new pricing systems afford the customers the luxury of specifying the expected planned style of usage. The usage can include the period of the day, expected number of hours, transactions and the expected number of access (see Figure 1).

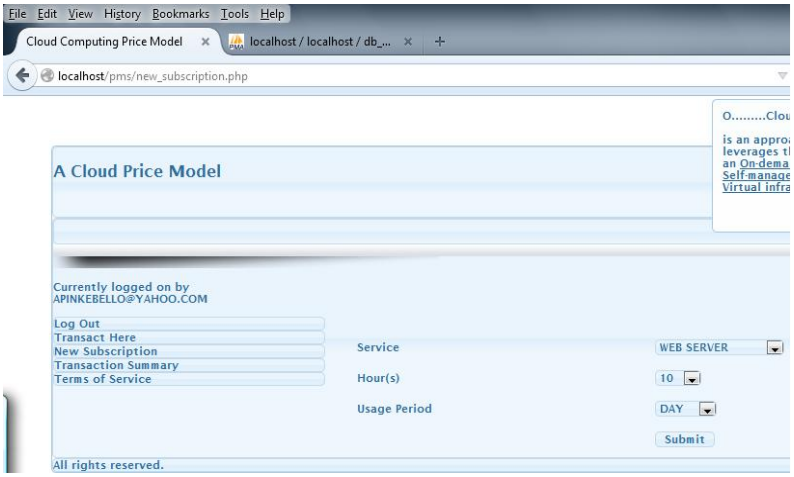


Figure-1: Usage Style Specification

The actual usage is compared with the expected usage specified earlier. The reduction display the discount given to the user due to the specifications made earlier. Violations can also be applied as a penalty in case there is a deviation from the expected usage earlier specified. The total amount due for payment will be computed based on the actual usage as against the flat price model earlier adopted by the SaaS (see Figure 2).

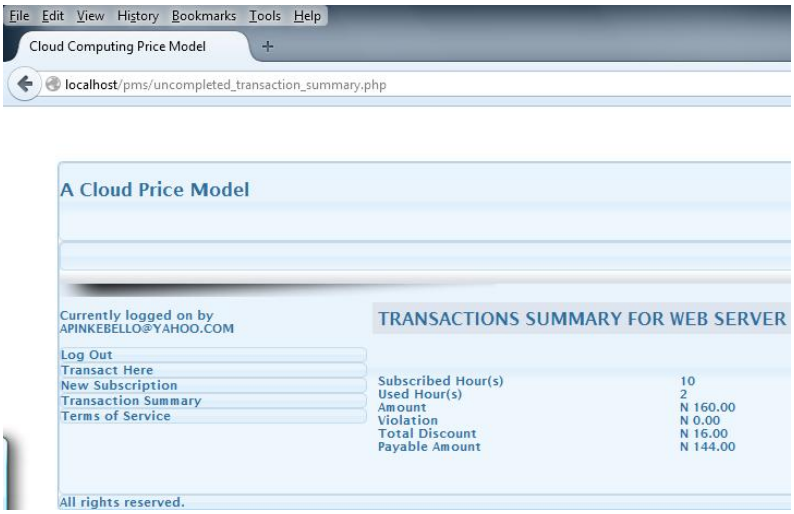


Figure-2: Actual Usage without Violation

If a customer predicts the usage of Cloud services, the Cloud provider could achieve a better utilization of the Cloud infrastructure and therefore can give the customer better prices.

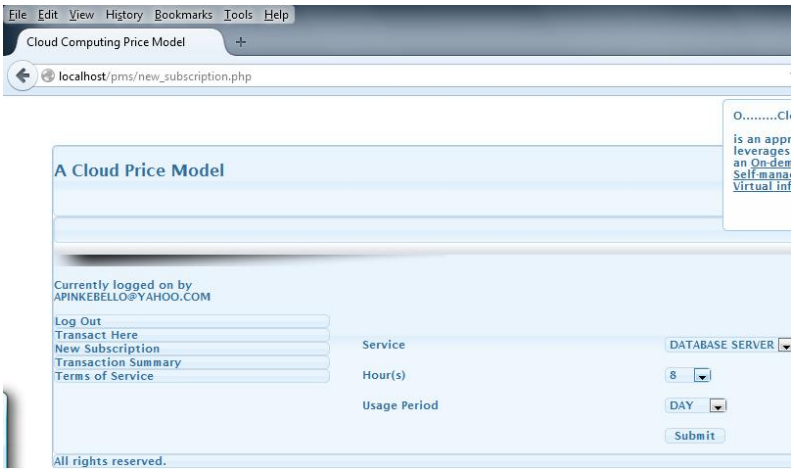


Figure-3: Initial User Specification

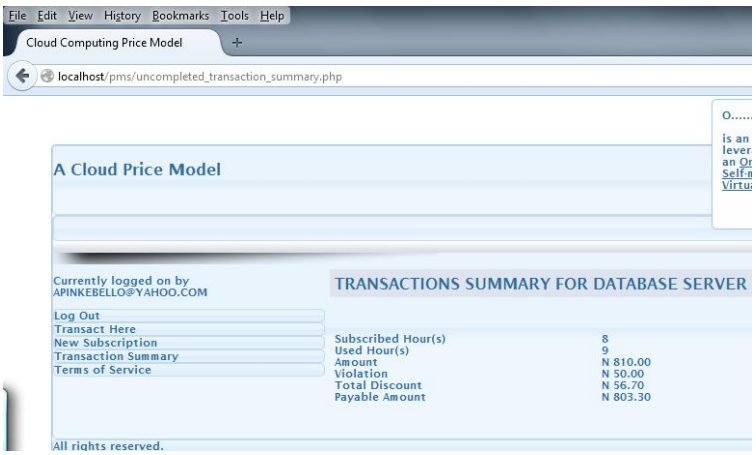


Figure-4: Actual Usage with Violation

6 Rewarding Models

A number of model can be used to reward users in addition to timing. This section discusses various rewarding price models which are common in business and could be experimented with cloud services.

6.1 Rewarding Loyal Customers

If a customer has been using a particular service for a period of time (e.g. several months), the customer is given a discount to reward loyalty. This implies the customer now pays less per e.g. number of users per month for this time period. This may retain the customer for the next time period. By this the SaaS provider may be assured of the customer's loyalty for the next time period. The customer might be retained for even a longer period, which is better for the provider because an existing customer is cheaper to maintain than new customers. Though a reduction in price as a reward for loyal customer may reduce the Providers income, but the income will still be higher than maintaining the static price and loosing customers.

6.2 Rewarding Night Usage

If users are given a significant discount for off peak usage like night period, say half the original price if the service is used between 10pm and 4am in the morning in addition normal day use. This may enable small software development companies to hire their developers for that period of the night. Since the developers may not necessarily be full time staff, so they are hired for this period, the job will get done. This offer is better given to existing users because they are already familiar with the service, hence may require fewer help desk attention because of the time of usage. The cloud user (Software Company) will be satisfied, it will spend less and the SaaS provider will also have a better utilization of the infrastructure.

6.3 Rewarding Transactions

Assume an online shop is hosted by a SaaS provider. If more users hit the service this increases the utilization of the service to the SaaS provider. The Cloud provider could give discount for a service that is well utilized. If increase in the actual transaction on a service can be rewarded, the SaaS provider may be able to generate more income. The online hop could subscribe to this, and will also introduce a sales offer for its client. Probably during winter that going out is not a welcome idea. The online shop gives a discount on some items to its client within a limited period of the day. It could add incentives like free delivery or buy two get three for online shopping. This may increase the number of transactions on the service as clients will most likely grab this opportunity. This will increase the number of clicks/transactions on this service to the online shop at that period. The online shop will have an increased income and the SaaS provider will also earn more.

6.4 Rewarding Increased Users

If a SaaS provider observe that one of its offer is getting low attention. The provider can allow probably a reduction in the trial period and give a discount. This may invite the attention of some SME companies that need such as service but has been hindered by the price attached. A software development company could hire many part time developers staff for a limited period that will enable the work to be finished as more people on the job will likely lead to a reduction in the finish up time. These will divert attention to the offer. Many more SMEs could be attracted by this the SaaS provider get more customer and more income.

All of the rewarding models described above can be combined, depending on the taste of the Cloud consumer. It can be seen that each of this new pricing generates more income for the provider and also gives flexibility to the Cloud consumer.

7 Conclusion

The proposed pricing model is an improvement over the existing flat Cloud pricing models for SaaS users. The possibility of employing differential pricing in Cloud services was re-enacted. Use case scenario was employed to illustrate possible implementation of differential pricing using timing. . Other possible ways of endearing users to cloud in form of rewards were also discussed. Cloud computing is still emerging, hence exploring the true utility nature will go a long way to endear users to it and increase its acceptability.

REFERENCES

[14]. "Twenty Experts Define Cloud Computing", SYS-CON Media Inc, http://cloudcomputing.sys-con.com/read/612375p.htm, Accessed June 18th 2012

[15]. Foster, I.; Yong Zhao; Raicu, I.; Lu, S.; , "Cloud Computing and Grid Computing 360-Degree Compared," Grid Computing Environments Workshop, 2008. GCE '08 , vol., no., pp.1-10, 12-16 Nov. 2008, doi: 10.1109/GCE.2008.4738445

[16]. Skilton, M. (2010). Building Return on Investment from Cloud Computing. White Paper. Cloud Business Artifacts Project, Cloud Computing Work Group, The Open Group.Http://www.opengroup.org/cloud/whitepapers/ccroi

[17]. Malone, T. W., Weill, P., Lai, R. K., D'Urso, V. T., Herman, G., Apel, T. G. & Woerner, S. L. (2006). Do Some Business Models Perform Better than Others? Working Paper 4615-06, Massachusetts Institute of Technology. Http://seeit.mit.edu/Publications/BusinessModelsPerformance12July2006.pdf. Accessed June 7th, 2012.

[18]. Strommen_Bakhtiar A. and Razavi A.R. (2011). Cloud Computing Business Models in Z. Mahmood and R. Hill (eds.) Cloud Computing for Enterprise Architectures, Computer Communications and Networks, DOI 10.1007/978_1_4471_2236_4_3, Springer Verlag London Limited 2011.

[19]. Ostenwalder, A., Pigneur, Y., and Tucci C.L., (2005). Clarifying Business Models: Origins, Present, and Future of the Concept. Communications of AIS, Volume 15, Article.

[20]. Magretta J. (2002), Why Business Models Matter, Harvard Business Rev., vol.80, no 5, 2002 pp 86-92

[21]. Fei Teng and Frederic Magoules (2009). Fundamentals of Grid Computing Theory and Algorithms and Technologies, chapter Future of Grid Resource Management, pages 133-153. Chapman & Hall/CRC.

[22]. Arto Ojala and Pasi Tyrvainen (2011) Developing Cloud Business Models: A Case Study on Cloud Gaming. IEEE Software. IEEE Computer Society. July/August pp 42-47

[23]. Victor Chang, Gary Wills and David De Roure (2010) A Review of Cloud Business Models and Sustainability. IEEE 3rd International Conference on Cloud Computing. DOI 10.1109/CLOUD.2010 **19 to**

[24]. Fei Teng and Frederic Magoules (2010) Resource Pricing and Equilibrium Allocation Policy in Cloud Computing. 10th IEEE International Conference on Computer and Information Technology (CIT 2010). DOI 10.1109/CIT.2010.70

[25]. Shadi Ibrahim, Bingsheng He and Hai Jin (2010) Towards Pay-As-You-Consume Cloud Computing 2011 1EEE International Conference on Services Computing. DOI 10.1109/SCC.2011.38

[26]. Iyer G.R, Miyazaki A.D., Grewal D. and Giordano M.," Linking web-based segmentation to pricing tactics" Journal of Product and Brand Management, 11 (5) (2002), pp. 288–302

[27]. Dolan, R. J., and Moon, Y. (2000). Pricing and market making on the Internet. Journal of Interactive Marketing, 14 (2), 56–73

[28]. Yuan-shuh Lii, Erin Sy, Internet differential pricing: Effects on consumer price perception, emotions, and behavioral responses, Computers in Human Behavior, Volume 25, Issue 3, May 2009, Pages 770-777, ISSN 0747-5632, http://dx.doi.org/10.1016/j.chb.2009.02.005

[29]. Wang H.Y , Jing Q.F., Chen R.S, He B.S., Qian Z.P., and Zhou L.D, (2010) "Distributed Systems Meet Economics: Pricing in the Cloud," Proc. Of the 2nd USENIX Workshop on Hot Topics in Cloud Computing (HotCloud'10) , Boston, USA

[30]. Weinhardt, C., Anandasivam, A., Blau, B., Borissov, N., Meinl, T., Michalk, W., Stößer, J.:Cloud computing–a classification, business models, and research directions. Bus. Info. Syst. Eng. 1, 391–399 (2009)

[31]. Rappa, M.A.: The utility business model and the future of computing services. IBM Syst. J. 43(1), 32–42 (2004)

[32]. Zhang, Q., Cheng, L. & Boutaba, R. (2010). Cloud Computing: state-of-the-art and research challenges. Journal of Internet Services and Applications, Vol. 1, No. 1, pp. 7—18.

[33]. Foster, I.; Yong Zhao; Raicu, I.; Lu, S.; , "Cloud Computing and Grid Computing 360-Degree Compared," Grid Computing Environments Workshop, 2008. GCE '08 , vol., no., pp.1-10, 12-16 Nov. 2008, doi: 10.1109/GCE.2008.4738445

[34]. Green M.A Third Generation Photovoltaics Advanced Solar Energy Conversion (Springer Verlag, Berlin , 2004)

[35]. Jaakko Jäätmaa, 2010 , Financial Aspects of Cloud Computing Business Models. Master's Thesis. Aalto University School of Economics.

[36]. Harmon, R., Demirkan, H., Hefley, B. & Auseklies, N. (2009). Pricing Strategies for Information Technology Services: A Value-Based Approach. Proceedings of 42nd Hawaii International Conference on System Sciences, pp. 1—10.

[37]. Paleologo, G. (2004). Price-at-Risk: A methodology for pricing utility computing services. IBM Systems Journal, Vol. 43, No. 1, pp. 20—31

[38]. Weinhardt, C., Anandasivam, A., Blau, B., Borissov, N., Meini, T., Michalk, W. & Stosser, J.(2009b). Cloud Computing - A Classification, Business Models, and Research Directions. Business Models & Information Systems Engineering, Vol. 1, No. 5, pp. 391— 399.

[39]. Amazon Web Services Simple Monthly Calculator, calculator.s3.amazonaws.com/calc5.html - Accessed June 18th 2012

[40]. Cloud Hosting Cost Calculator, GoGrid Cloud Hosting, www.gogrid.com/.../cloud-hosting-pricing-calcu...Accessed June 18th 2012

[41]. Computing TCO Calculator www.rightscale.com/tco-calculator, Accessed June 18th 2012

[42]. Sales Force Pricing and Editions www.salesforce.com/crm/editions-pricing.jsp, Accessed June 18 2012

TNC Transactions on Networks and Communications
SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
VOLUME 2, ISSUE 5
ISSN: 2054 - 7420

An Organizational Role-based Extrusion Detection Model with Profile Migration

Tirthankar Ghosh, Kasun Abeykoon and Thusith Abeykoon
Department of Computer Science and Information Technology
College of Science and Engineering
St. Cloud State University
tghosh@stcloudstate.edu

ABSTRACT

Intrusion detection and prevention systems play a crucial role in the overall information security implementation of today's organizations. Traditionally, signature-based and anomaly-based detections have been the two main methods of detection and prevention techniques. Signature-based intrusion detection systems are excellent in detection and performance, but they are vulnerable to unknown threats like zero-day attacks. Extensive research have been conducted on anomaly detection and prevention based on users' behavior profiling. However, as insider attacks increase, it has become equally important to monitor and analyze extrusion attempts. Behavior-based profile creation has a promising future in extrusion monitoring. However, profiling individual behavior has its limitations in that it tends to incorporate unintended behavior into the normal profile. In this study, user's organizational role has been integrated into profile creation further reducing number of false positives. A prototype of the model is tested with three users belonging to three different roles. A profile migration scheme is proposed to import user profiles at various login locations.

Keywords – Extrusion detection, Role-based profile modeling, Profile migration

1 Introduction

For many years, security experts have given much priority to increase network perimeter security because most attacks were targeted on breaching perimeter security. But with the development of technologies like mobile computing, cloud computing, distributed processing and increasing reliance on web-based applications, cyber-attacks are becoming more and more complex and advanced with time. Even though it is extremely important to protect the networks from external attacks, internal attacks can do far more damage as these attacks come from trusted insiders.

Attacks do not always occur in the same pattern. Attackers will try different kinds of methods to achieve their goal. Intrusion Detection and Prevention Systems (IDPSs) traditionally use signature-based approach, which is a very effective method and achieves better performance in detecting known threats; however they are ineffective in detecting new and unknown attacks. To overcome this, researchers have developed anomaly based IDPSs, which use behavior analysis to detect intrusion. Using anomaly based detection, IDPSs compare current system, user, or network status with a previously

DOI: 10.14738/tnc.25.473
Publication Date: 23th September 2014
URL: http://dx.doi.org/10.14738/tnc.25.473

created profile to detect anomalous behaviors. The main drawbacks in anomaly-based detection methods are managing large number of profiles, increasing false negatives and false positives.

Anomaly based IDPSs create user profiles to identify each user in the system. These systems consist of a learning phase and a detection phase. In the learning phase they gather data from each user to create profiles, and in the detection phase they compare users' data in real time with previously gathered information stored in the profiles. The main drawback in this process is that there is no method to validate actions done by the users in the learning process. Therefore, unauthorized actions might be added to the users' profiles. To overcome this issue, the users' organizational role based on their job function gets introduced to the anomaly profile creation process. A group profile is created for each role by analyzing users' data within a specific group. Since there are standard authorized activities a user can do in a specific job role, the individual unauthorized activities will be ignored during profile creation. A good framework to build anomaly profiles for large organizations was introduced by role based profile analysis [1] which use both role based profiles and individual profiles to detect attacks. This research has extended their model in creating combined role-based and individual-based profiles by designing and implementing an algorithm to detect deviations.

The authors of this paper have also presented a prototype to migrate user profiles at multiple login locations. Most of the previous work related to IDPS user profiles have been done to create those profiles, and not much research is present in the literature involving their transferability. The need of designing a scheme to migrate user profiles within an organization has also been addressed in this research.

The rest of the paper is organized as follows. Section 2 discusses some relevant literature in the context of anomaly detection and profile migration, section 3 describes our design and prototype implementation, section 4 discusses the results, and section 5 describes our profile migration scheme. Finally, section 6 concludes the paper.

2 Literature Review

Research in anomaly based profile creation is a rapidly developing area in the information security field. Researchers use a wide range of methodologies to develop profiles to obtain high accuracy detection. One of the main issues in maintaining user profiles is how to handle users' behavior change over time. This is also known as Concept Drift [2], and anomaly profiles should be up-to-date to detect the new user behavior. A dynamic normal profiling system [3] was introduced to solve this problem and the special algorithms SVM (Support Vector Machine) and FLORA 2 (Floating Rough Approximation) were used to update user profile during extended period of time. SVM algorithm creates normal profiles by filtering irrelevant data, and FLORA2 uses a dynamic sliding window to update patterns with time. Multiple classifiers were used to improve performance in anomaly-based detection by using the Kyoto 2006+ dataset obtained from honeypots at the Kyoto University [4]. A Multi-Level Intrusion Detection System (ML-IDS) was proposed in [8] to use three parameters to detect intrusions: traffic flow, packet header, and payload. Instead of obtaining results from individual parameters, ML-IDS used a decision fusion technique to get better results. Another study was conducted in [5] by observing number of running applications, number of open windows, application performance, websites viewed, and keystroke analysis.

Further research have shown that intrusion detection on encrypted sessions is difficult because the system has to decrypt it before analyzing the data. Protomon [6] is one of the unique anomaly based IDS which was developed to detect malicious behaviors in cryptographic and application level protocols in encrypted sessions. Casual Relation Minor framework (CR-Miner) was developed to identify anomalous events on a host [9], by discovering the relations between user activities and networks traffic.

Since Role Based Access Control (RBAC) [7] was introduced, IDPSs have integrated it to get better performance and accurate detection. The advantages of role based access control along with scalable administration, separation of duties and least privilege were all taken into account when developing the Composite Role Based Monitoring (CRBM) system [10]. The main advantage of CRBM is it maps vertically and horizontally between different levels and domains. CRBM also used user's organizational role, applications, and operating systems in its behavior monitoring system. One of the other main technologies integrating RBAC is database management systems. Database traces stored in log files were used, to model user's normal behavior and identify intrusion [11]. Furthermore, they have integrated RBAC, which will reduce the number of user profiles to compare, and alert when a user tries to access system resources beyond the user's role permission.

Anomaly based detection systems use statistical analysis to compare current user data with predefined user profiles. Standard deviation is a commonly used measure to detect deviation. For more advanced comparisons, the Kullback-Leibler (K-L) divergence [12] has been used. K-L divergence was used to detect anomalous behaviors in wireless communication signals [13] from interferences and to analyze DNS traffic for Domain-Flux attacks from Botnets [14].

There has not been much discussion in the literature on designing an efficient communication framework for importing user profiles in an intrusion detection environment. Earlier studies suggested a communication mechanism for cooperation between multiple intrusion detection agents using a new protocol named Information Exchange Protocol to avoid limitations of intrusion detection system with a controlling center [15]. Information exchange protocol use UDP datagrams for its communications. It also uses four kinds of event formats in exchanging messages as event receiving, event sending, announcement receiving and announcement sending. The Intrusion Detection Working Group (IDWG) suggested a protocol called Intrusion Detection Exchange Protocol (IDXP) to enable communication between intrusion detection systems [16] over a connection oriented protocol. Also, IDXP is an application-level protocol which supports confidentiality, integrity and mutual authentication between intrusion detection systems. IDXP supports exchange of data in a structured or unstructured format. Structured data includes messages in a format defined in Intrusion Detection Message Exchange Format (IDMEF), while unstructured data includes unstructured text and binary data.

All of these protocols work to exchange data between intrusion detection systems, not between other systems. Therefore, the importance of a data exchange mechanism between intrusion detection systems and other devices such as an authentication server is growing.

3 Design and Implementation

The basic premise of the model is to integrate the role-based profile with the individual-based profile. Four parameters are chosen to represent the profiles, two for each profile: CPU and memory utilization for the individual profile, and the number of processes and network connections for the role-based profile. The deviation is measured by the Kullback-Leibler (K-L) divergence as given below in Equation 1.

The Kullback-Leibler divergence is a statistical function used to measure the proximity of two discrete probability distributions. If P and Q are two probability distributions, K-L divergence is give by Equation 1 below:

$$D_{KL}(P||Q) = \sum_i \ln(\frac{P(i)}{Q(i)})P(i) \quad (1)$$

K-L divergence satisfies the following properties:

- *Is always a positive number*
- *Is equal to zero if P(i) = Q(i) (If two distributions are identical)*
- *Is not defined when P(i) ≠ 0 and Q(i) = 0*

Instead of using Standard Deviation (SD), which measures how much the new data set is deviated from the mean, K-L divergence is used in this study to get a more precise comparison with user profiles and current data. A value of 0.5 is added to the values before calculating the K-L value to eliminate P(i) ≠ 0 and Q(i) = 0 [13].

Our model is divided into three sections, the learning phase, the profile creation phase, and the detection phase. The learning phase describes how the data is gathered and saved to create user profiles. The profile creation phase describes how individual and role-based baseline profiles are created. Lastly, the detection phase detects intrusions in real time by measuring deviation from baseline profiles. Each phase is described in details below.

3.1 Learning phase

In the learning phase the system collects user's CPU usage and memory usage (for individual profiles), and number of running processes and number of established network connections (for role-based profiles) every five seconds. These data are grouped into one-minute sections to find the average values for every minute and every third minute. The flowchart is shown in Figure 1.

Each value will be recorded and stored in a file for use during the profile creation phase.

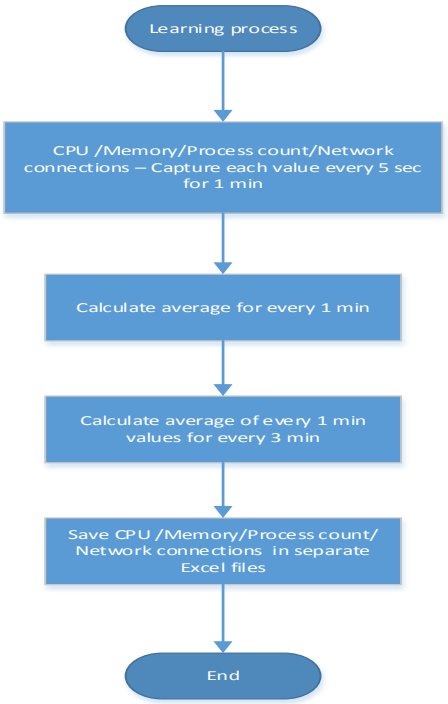


Figure 1 Learning phase flow chart

3.2 Profile creation phase

In the profile creation phase, the system reads each user's CPU and memory data and creates an individual profile for each user. Simultaneously, the system collects all users' process and network usage data based on their organizational role. Separate role-based profiles are created for each role. The flowchart is shown in Figure 2.

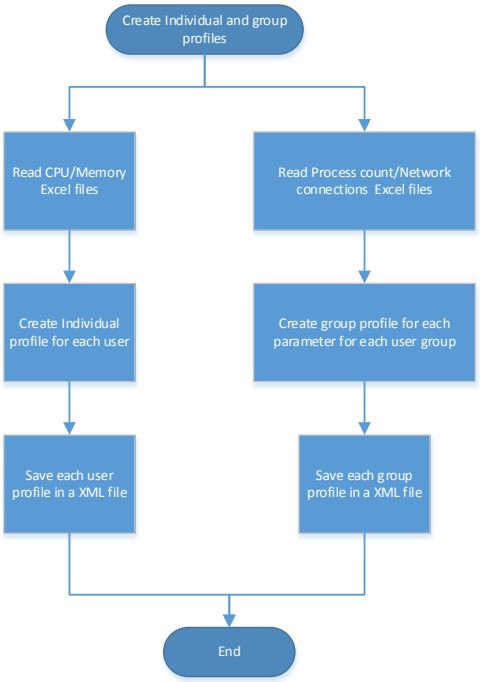


Figure 2 Profile creation phase flow chart

3.3 Detection phase

In the detection phase, before running the actual detection process, a trial detection process will run, which includes several tests to find the limits and thresholds of K-L values. The detection phase is divided into three sub phases as described below.

In the first phase, the system reads the user's individual and role-based profile data and stores them in memory. Then the system starts and captures the user's real-time CPU usage, memory usage, number of running processes, and the number of established network connections for every five seconds. While the system is computing K-L values by analyzing this data, several test tools are used to generate high CPU activity along with memory, process and network utilizations. This method will help to define what the lowest and highest K-L values are when the system is in normal use. It also shows how K-L values increase with high CPU, memory, process and network utilizations, and finally, to define the threshold K-L values for each parameter. The overall algorithm is shown in Figure 3.

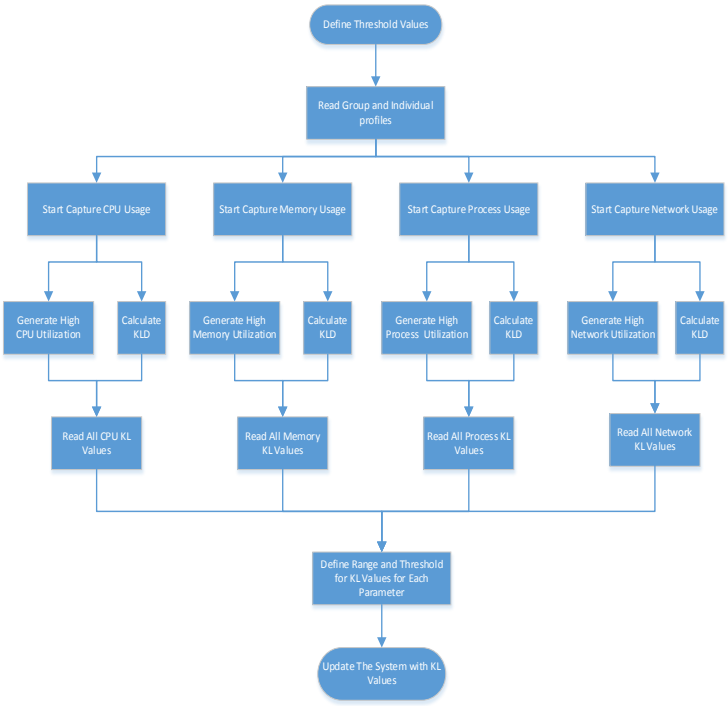


Figure 3 First detection phase flow chart

Before starting the first phase, the minimum utilization, maximum utilization, and intervals between them are defined for each parameter. When the process starts to capture usage of each parameter, the test tools which are used to simulate high utilization also starts from minimum utilization to a maximum utilization. If the normal system utilization is higher than the predefined minimum utilization, the test tool will start to simulate from normal system utilization. All these values are recorded in a file. At the end of this process, for each parameter, the K-L values when the system runs in normal behavior and in intervals between normal and maximum system utilizations are identified and inserted into the system to run in the next phase. The K-L value when the system runs in normal utilization will be the baseline K-

L value for each parameter, and the K-L values for intervals above the baseline will be used as the threshold values.

In the second phase, the system loads the user profiles first and starts to capture user data. At the end of each minute, the system compares data with the profile and then computes the K-L value and compares it with the current threshold K-L value. All observations during normal system behavior with K-L values greater that this threshold are flagged as false positives; all observations under stressed behavior with K-L values lower than the threshold are flagged as false negatives. False positives and negatives are plotted with increasing threshold values to find the crossover point. The algorithm is shown as a flowchart in Figure 4.

In the final phase, the threshold values are added systematically from the second phase. Then the system again starts to capture the user's real time data to compute K-L values. The inputs are passed through the individual profile-matching engine to analyze their behavior and measure deviation. If there are any deviations from the predefined threshold, output from the individual profile-matching engine are passed onto the role-based profile-matching engine. If deviation is detected in the second phase, the incident is marked as anomalous. The algorithm is shown in Figure 5. The rationale for using individual profile-matching engine before role-based profile-matching engine is as follows: if the role-based profile is checked first with the number of connections and number of processes running, a deviation may be passed onto the second stage with a higher number of process running, which may not necessarily mean that CPU and memory utilization will be higher. That will not be flagged as a deviation at the second stage. On the other hand, if the individual profile is checked first with the CPU and memory utilization, any deviation with high utilization will have a higher chance of getting flagged as a deviation at the second stage, as higher utilization may have been caused by more processes or more connections, thus reducing overall false negatives. In this phase, the system knows the K-L values for normal and suspicious usage, and from the K-L output of user, the system can define whether the current behavior is suspicious or not.

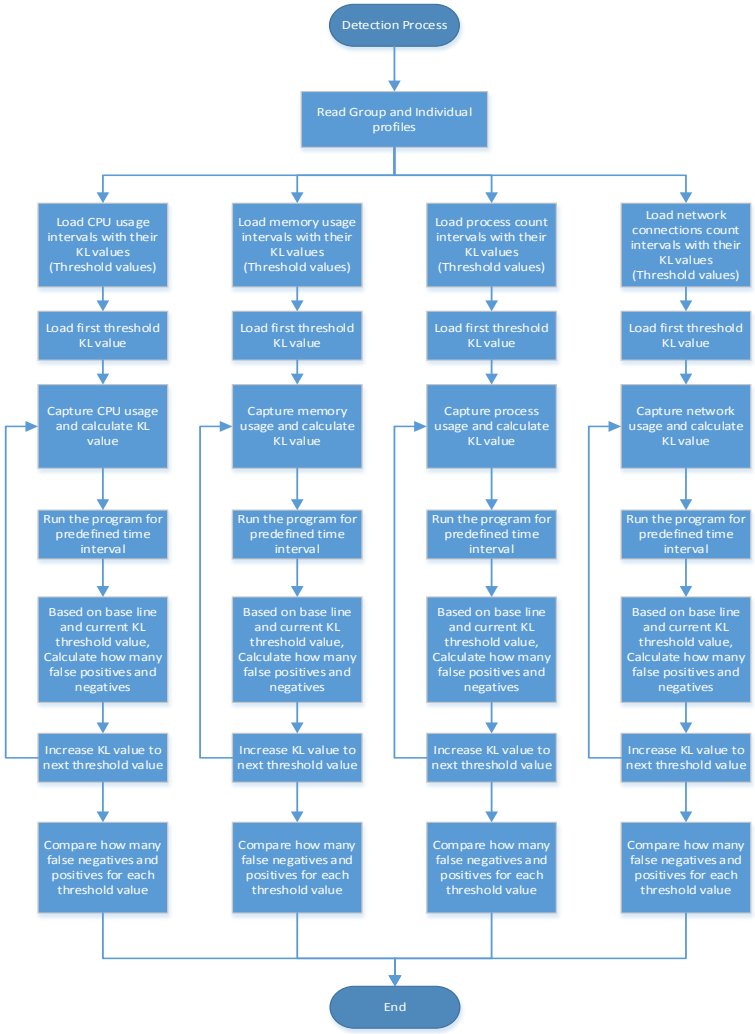


Figure 4. Second detection phase flow chart

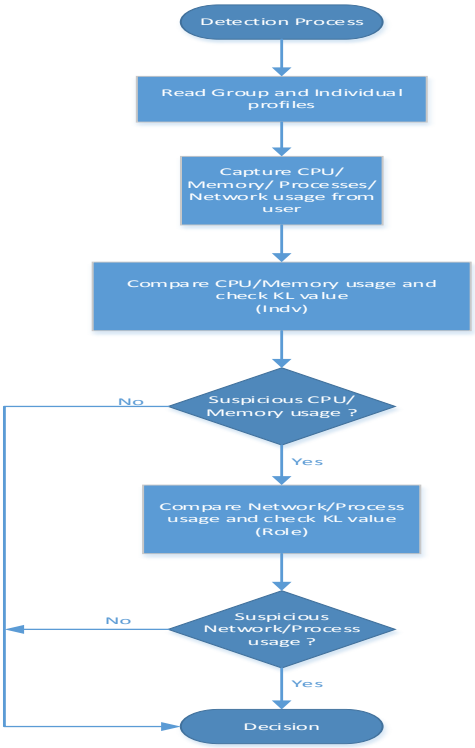


Figure 5. Final detection phase flow chart

4 Discussion of Results

Python 2.7 was used as the primary programming language to collect data for creating individual and role-based profiles. These profiles were saved as XML documents. Three users were selected belonging to different roles, and asked to run the programs in order to capture their CPU activity, memory usage, as well as process and network usages. Data was collected for three weeks in a supervised environment, and each user's individual and role-based profiles were created. One challenge was to determine the appropriate threshold value for using in the detection phase. We stressed the systems by injecting high CPU, memory, network connections, and processes, and recorded the parameters. Based on whether each value is between the baseline and threshold value or above the threshold value, the outcomes are defined as either a false negative or a false positive respectively. Plotting false positives and negatives with increasing threshold, we were able to find the optimum point, and use that as our predefined threshold. The tests ran for several days to capture user data and were grouped by each day.

4.1 Individual Profiles

Figures. 6, 7, and 8 show each user's CPU usage. For the first user, the normal CPU usage is 10% or below, and the maximum usage is generally 50%. The user's threshold K-L value is between 10-20%. The optimal K-L value is 0.6, and the relative CPU usage is 15.69%.

For the second user, the normal CPU usage is 20% or below, and the maximum usage is generally 50%. The user's threshold K-L value is between 30-40%. The optimal K-L value is 0.92, and the relative CPU usage is 32.09%.

For the third user, the normal CPU usage is 30% or below, and the maximum usage is generally 90%. The user's threshold K-L value is between 40-50%. The optimal K-L value is 0.45, and the relative CPU usage is 44.47%.

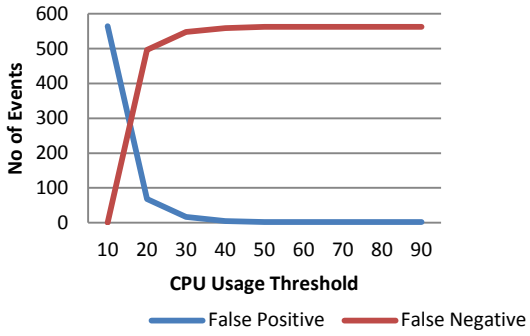


Figure 6. Optimum threshold for the first user's CPU usage

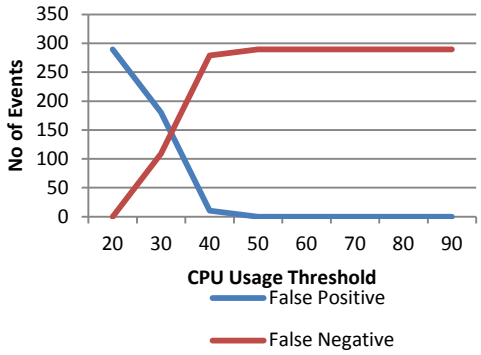


Figure 7. Optimum threshold for the second user's CPU usage

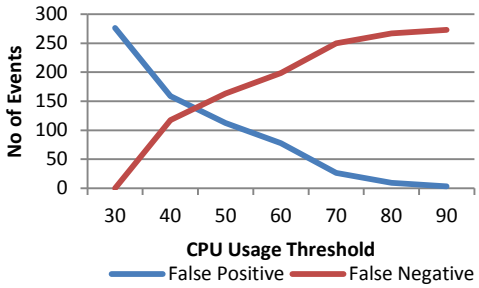


Figure 8. Optimum threshold for the third user's CPU usage

Figures 9, 10, and 11 show each user's memory usage. For the first user, the normal memory usage is 60% or below, and the maximum usage is generally 90%. The user's threshold K-L value is between 60-70%. The optimal K-L value is 1.56, and the relative memory usage is 65.69%.

For the second user, the normal memory usage is 70% or below, and the maximum usage is generally 80%. The user's threshold K-L value is between 70-80%. The optimal K-L value is 1.275, and the relative memory usage is 75%.

For the third user, the normal memory usage is 50% or below, and the maximum usage is generally 90%. The user's threshold K-L value is between 60-70%. The optimal K-L value is 2.0, and the relative memory usage is 66.9%.

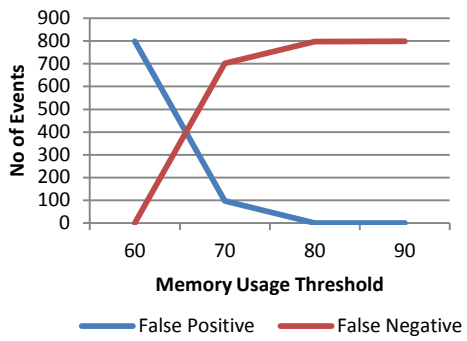


Figure 9. Optimum threshold for the first user's memory usage

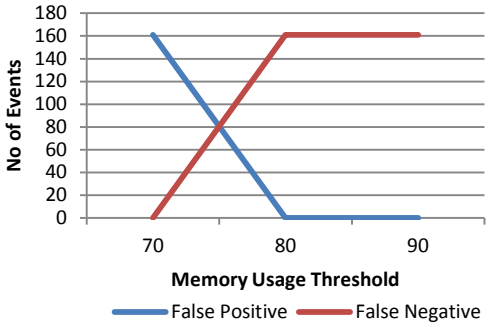


Figure 10. Optimum threshold for the second user's memory usage

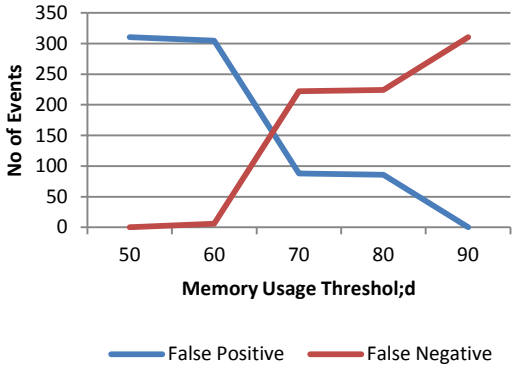


Figure 11. Optimum threshold for the third user's memory usage

4.2 Role-based Profiles

Role-based profiles show each user's process (application) and network connections usage, where each user is selected to represent a different organizational role.

Figs. 12, 13, and 14 show each group's (role) process usage. The normal process usage for the first group is 120 processes or below, and the maximum usage is generally 130 processes. The user's threshold K-L value is between 120 and 130 processes. The optimal K-L value is 1.362, and the relative application usage is 125 applications.

The normal process usage for the second group is 80 processes or below, and the maximum usage is generally between 100 to 150 processes. The user's threshold K-L value is between 90 and 100 processes. The optimal K-L value is 1.4, and the relative application usage is 97 applications.

The normal process usage for the third group is 80 processes or below, and the maximum usage is generally 100 processes. The user's threshold K-L value is between 80 and 90 processes. The optimal K-L value is 1.362, and the relative application usage is 85 applications.

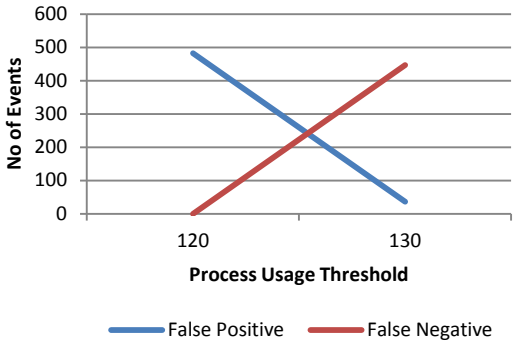


Figure 12. Optimum threshold for the first group's process usage

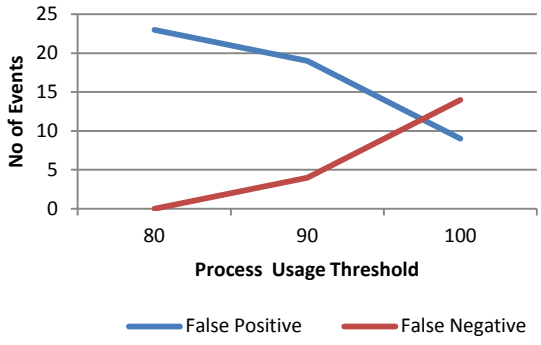


Figure 13. Optimum threshold for the second group's process usage

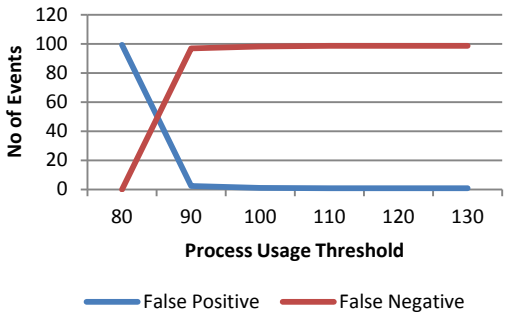


Figure 14. Optimum threshold for the third group's process usage

Figures 15, 16, and 17 show each group's (role) network usage. The normal network usage for the first group is 30 connections or below, and the maximum usage is generally 120 connections. The user's threshold K-L value is between 40 and 50 connections. The optimal K-L value is 2.4, and the relative network usage is 41 network connections.

The normal network usage for the second group is 20 connections or below, and the maximum usage is generally 140 connections. The user's threshold K-L value is between 20 and 40 connections. The optimal K-L value is 2.49, and the relative network usage is 38 network connections.

The normal network usage for the third group is 10 connections or below, and the maximum usage is generally 90 connections. The user's threshold K-L value is between 30 and 40 connections. The optimal K-L value is 1.7, and the relative network usage is 33 network connections.

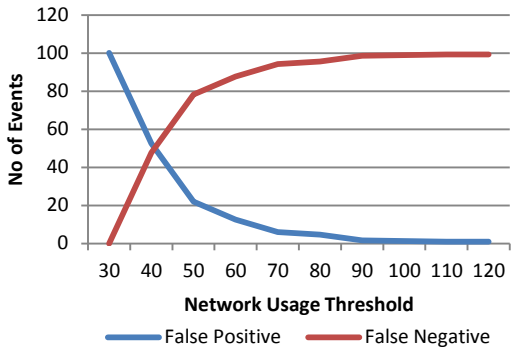


Figure 15. Optimum threshold for the first group's network usage

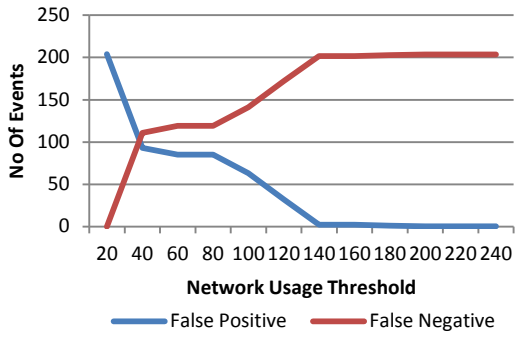


Figure 16. Optimum threshold for the second group's network usage

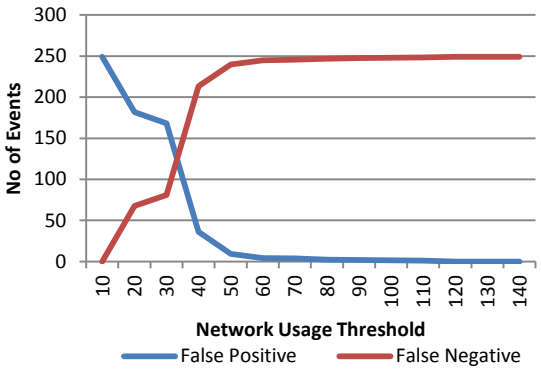


Figure 17. Optimum threshold for the third group's network usage

4.3 Final detection and false positive analysis

Our algorithm uses the individual profile-matching engine before the role-based profile-matching engine (rationale already given in section 3.3). In this section we analyze false positives, and show that our approach succeeds in reducing false positives. The analysis is done on data collected during the detection phase with real-time data under supervised condition with no malicious activity.

Table 1 summarizes results for the first user. It can be seen that about 14% of events were flagged as intrusion in the first stage by the individual-based detection engine, while only 4% of events were flagged as intrusion by the system. This significantly reduces false positives. Similarly, tables 2 and 3 summarize results for the second and third users respectively. Flagged events were reduced from 18% to null for the second user, and from 53% to 3.5% for the third user. These tests show that our system is capable of significantly reducing false positives.

Table 1. False Positive Analysis for the First User

	No of Events	Percentage
Total	2453	
Events passed Individual profile test	2104	85.70%
Events failed Individual profile test but passed group profile test	250	10.19%
Events failed both tests	100	4.07%

Table 2. False Positive Analysis for the Second User

	No of Events	Percentage
Total	1078	
Events passed Individual profile test	878	81.44%
Events failed Individual profile test but passed group profile test	200	18.55%
Events failed both tests	0	0%

Table 3. False Positive Analysis for the Third User

	No of Events	Percentage
Total	958	
Events passed Individual profile test	451	47.07%
Events failed Individual profile test but passed group profile test	473	49.37%
Events failed both tests	34	3.54%

5 Profile Migration Scheme

There is a need to migrate user and group profiles, as users are becoming mobile and login to the domain from various locations. We have proposed such a scheme to migrate these profiles. Also to move the user profiles from one location to another, they have to be saved in one or more locations where they can be accessed from any host within the organizational network. These locations should act like data-stores, which could house all user profiles in specific format and profiles should be identified according to the user credentials.

Our design has three main components: client, authentication server, and data store. Client is the user entry point to the organizational network. Mainly this would be a computer where users could log into

the organization network and their credentials will be authenticated by an authentication server. After the profiles are moved to the client machine, they are being used by the IDPS system residing in the client machine.

The authentication server is the main device that glues the system togeather. It communicates with both client and the data store, and transfers profiles to the client. It also authenticates each user logging into the host machine. The main tasks of this server are to authenticate the user, apply any policies related to the user, and transfer the user profiles (both individual and role-based) from the data store to the client.

The data store is the repository of IDPS user profiles. These user profiles are indexed according to any user credentials such as user name or user id. Each user profile will be saved as an XML file to be supported by the data store. It would be an added advantage if the data store supports any regular expression based searches within the data store. The data store only communicates with the authentication server to transfer profiles.

The process flow is shown in Figure 18 below. Users are assumed to connect to the organization's network domain for authentication by the domain controller. The process starts from the user. Initially, the user provides login credentials to the client machine. The client machine initiates the login process, and provides the supplied credentials to the authentication server. On success, the authentication server initiates the profile transfer process to identify and transfer the IDPS individual and role-based profiles to the client machine. After the user related processes are done, the user login process will complete its task and the user can access the machine.

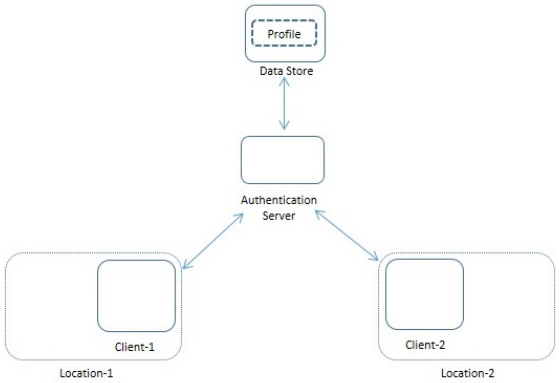


Figure 18. Process flow for user profile migration scheme

The design was implemented using Windows Server 2003 as the directory server (domain controller), and Microsoft SQL Server 2008 as the data store, all running on virtual hypervisors. VMWare Workstation was used as the virtual hypervisor. Selecting a platform for a data store was based on several considerations. First was a directory server, second was a database server, and third was storing the profiles in a file system. Storing the profiles in a directory server was not considered because of two reasons; one, it would add an extra load to the current directory servers; two, adding the profile data in a seperate system will make the data store an isolated system which can be managed and controlled seperately without entering to the directory server. Selecting a database as a data store rather than a

flat file system or directory service provides several advantages such as easier indexing, iplementing security, and open to sql queries for easier data searching and manipulation.

One of the initial requirements of the design is to transfer the profiles which are saved in the data store to the host machine. First, the profile requesting component, which is the client or the server, has to identify the logon process and initiate a communication channel to the data store to get the profiles. Then the data store has to be queried to get the specific users' profiles. Since the profiles are indexed according to user credentials such as user name or user id, the initiating component has to have the requesting user details with the query. A successful query will identify the profiles (both individual as well as role-based) for the user which will be transferred to the client machine to be used by the IDPS. Since the authentication server runs as a directory server, it has the ability to identify the logon process for a user. Also since the data store works as a database, a database client tool will help to connect the datastore and run a query to get the profiles. The client tool should have the user's id to get the profiles. To complete this process a Visual Basic script was written to connect to the database and get the profiles. This script stays in the "sysvol" directory on the directory server and runs on the client machine when the user logs on. The script is executed as a part of the user logon process, and executes the database client tool to connect to the data store and query the user's profiles. If found, the script gets the profiles and transfers them on to the client machine.

6 Conclusion and Future Work

As insider attacks are growing, anomaly-based extrusion detection systems are becoming popular in many organizations. In this research we have designed such a system based on integration of individual profiles and organizational role-based profiles. We have implemented a prototype of the system, and demonstrated that our system is capable of reducing false positives significantly. We have also implemented a profile migration scheme, which helps to migrate users IDPS profiles to various locations within and outside the organizational boundary as long as the users log in to the domain.

A future extension of this research will be to make the user profiles (both individual and role-based) dynamic based on a feedback mechanism. Adding parameters such as user's geographical location and devices used will also help in building better profiles.

REFERENCES

[1]. Park, J. S., & Giordano, J., "Role-Based Profile Analysis for Scalable and Accurate Insider-Anomaly Detection", *25th IEEE International Performance, Computing, and Communications Conference*, pp. 463-470, 2006.

[2]. Wang, S., Schlobach, S. & Klein, M. C. A., "What Is Concept Drift and How to Measure It", In P. Cimiano & H. S. Pinto (eds.), *EKAW* , pp. 241-256, : Springer, 2010.

[3]. Kim, J. Y., Gantenbein, R. E., & Sung, C. O., "Dynamic Normal Profiling for Anomaly Detection Systems", in *Proc. 3rd Int. 3rd International Conference on Convergence Technology and Information Convergence*, pp. 27-32, 2008.

[4]. Kishimoto, K., Yamaki, H., & Takakura, H., "Improving Performance of Anomaly-based IDS by Combining Multiple Classifiers", in *Proc. IEEE/IPSJ International Symposium on Applications and the Internet*, pp. 366-371, IEEE Computer Society, 2011.

[5]. Pannell, G., & Ashman, H., "Anomaly Detection over User Profiles for Intrusion Detection", *Australian Information Security Management Conference,* Perth, Australia, 2010.

[6]. Joglekar, S. P., & Tate, S. R., "ProtoMon: Embedded Monitors for Cryptographic Protocol Intrusion Detection and Prevention", in *Proc. International Conference on Information Technology: Coding and Computing,* pp. 81- 88, 2004.

[7]. Ferraiolo, D. F., & Kuhn, D. R., "Role-Based Access Controls", in *Proc. 15th National Computer Security Conference*, pp. 554 - 563, 1992.

[8]. Al-Nashif, Y., Kumar, A. A., Hariri, S., Luo, Y., Szidarovsky, F., & Qu, G., "Multi-Level Intrusion Detection System (ML-IDS)", in *Proc. International Conference on Autonomic Computing*, pp. 131-140, 2008.

[9]. Zhang, H., Banick, W., Yao, D., & Ramakrishnan, N., "User Intention-Based Traffic Dependence Analysis for Anomaly Detection", *in Proc. IEEE Symposium on Security and Privacy Workshop*, pp. 104-112, 2012.

[10]. Park, J. S., & Ho, S. M., "Composite Role-Based Monitoring (CRBM) for Countering Insider Threats", in *Proc. Second Symposium on Intelligence and Security Informatics*, pp. 201-213, Heidelberg, Germany, 2004.

[11]. Kamra, A., Terzi, E., & Bertino, E., "Detecting Anomalous Access Patterns in Relational Databases", *The VLDB Journal — The International Journal on Very Large Data Bases*, vol 17(5), pp. 1063-1077, August 2008.

[12]. Kullback, S., Leibler, R. A., "On Information and Sufficiency", *The Annals of Mathematical Statistics*, vol 22(1), pp. 79-86, March 1951.

[13]. Afgani, M., Sinanovic, S., & Haas, H., "Anomaly Detection using the Kullback-Leibler Divergence Metric", in *Proc of the 1st International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL 08), 2008.*

[14]. Yadav, S., Reddy, A. K. K., Reddy, A. L. N. & Ranjan, S., "Detecting Algorithmically Generated Domain-Flux Attacks With DNS Traffic Analysis", *IEEE/ACM Trans. Netw.*, 20, pp. 1663-1677, 2012.

[15]. Zhang , W., Wang , L., Fu , X., & Teng, S., "Research on Communication Mechanism Among Cooperating Multi-Intrusion Detection Agents", *ICCI '06 Proceedings of the 2006 5th IEEE International Conference on Cognitive Informatics*, *pp.* 743-748, 2006.

[16]. Feinstein, B., & Matthews, G., "RFC4767 *The Intrusion Detection Exchange Protocol (IDXP)*", retrieved from http://www.ietf.org/rfc/rfc4767.txt, March 2007.

TNC Transactions on Networks and Communications
SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
VOLUME 2, ISSUE 5
ISSN: 2054 - 7420

Rate Dynamics and Transmission Algorithms in Clustered Sensor Networks

[1]A.T. Burrel [2]P. Papantoni-Kazakos
[1]Dept. of Computer Science, Oklahoma State University, USA
[2]Dept. of Electrical Engineering, University of Colorado Denver, USA
tburrell@okstate.edu; titsa.papantoni@ucdenver.edu

ABSTRACT

We consider wireless clustered sensor networks consisting of large numbers of life-limited sensors, where the size of each cluster is determined by the data rate generated by the sensors it contains. The limited sensors' life-span and their possible mobility induce variations in the sensor population and data rate dynamics. The variations of the sensors' population necessitate the deployment of random access data transmission algorithms, where stable such algorithms lie within the limited sensing class. We propose the deployment of the Limited Sensing Stack Random Access Algorithms (LSSRAAs), with a modification which allows the accommodation of high priority data. We evaluate the performance of the LSSRAAs in the presence of sensor expirations and recommend design specifications for systems with low versus relatively high energy reserves.

Keywords: Clustered Sensor Networks; dynamic data rates; limited sensor life span; limited sensing random access transmission algorithms; priority data; dynamic cluster reconfiguration.

1 Introduction

We consider wireless sensor networks containing large numbers of possibly mobile sensors, normally distributed over a wide geographical area. Such networks are then comprised of sensor clusters, where the size of each cluster may be determined by the communication range of the devices, in conjunction with the aggregate data rate generated by the sensors contained in the cluster. The clusters are generally connected via a backbone network which also normally includes a base station that performs global data processing operations.

The distinguishing feature in wireless sensor networks is the limited life-spans of the sensors, as induced by their energy consumption. Interesting results focusing on energy consumption have been obtained by several researchers: Bounds on energy conservation techniques have been derived in [1], role assignments targeting energy conservation have been developed in [2], energy conservation routing techniques have been proposed in [3], [4], [5] and issues arising due to energy conservation have been discussed in [5], [6]. In addition, topological and node-cooperation issues have been included in [7], [8], [9], while approaches to performance monitoring have been presented in [10]. An interesting rate allocation algorithm has been presented in [11], which is based on a modification of the max-min routing in [12] and the lexicographic linear programming approach in [13]. In [14], a dynamic rate

DOI: 10.14738/tnc.25.482
Publication Date: 23th September 2014
URL: http://dx.doi.org/10.14738/tnc.25.482

allocation methodology has been presented that is facilitated by the powerful data rate monitoring algorithm in [15], [16],[17], while an architectural reconfiguration approach based on data rate monitoring is presented in [18] and [19]. In [20], energy efficient clustering algorithms are proposed, including a discussion on LEACH approaches, where energy consumption is assumed to be strictly a function of geographical distances and where transmission collisions are completely ignored. Considering transmission protocols, those proposed or partially implemented (in Zigbee, IEEE802.15.4 etc.) within the Random Access class are ALOHA-based and are characterized by well-known instabilities that pull the throughput down rapidly to zero, as the user population increases.

Wireless sensor networks are designed to satisfy signal processing objectives. Thus, their performance metrics are determined by those of the latter objectives [21], [22]. When time constraints are imposed on high accuracy signal processing operations, the consequence is increased required overall data rates. At the same time, due to sensor expirations and subsequent time-varying sensor populations, the data transmitting algorithms deployed by the sensors must be within the random access class. In particular, for stability and implementability, the class of Limited Sensing Random Access Algoriths (LSRAAs) must be then deployed by the sensors in each cluster. Such are the algorithms in [23], [24], [25], [26],whose performance is a function of the their input data rates, while, at the same time, the life-span of the sensors is a function of the data rates generated within the clusters containing them, [27], [3], [4], [11]. Thus, required overall data rates, in conjunction with rate-dependent LSRAA performance and sensor life-spans, necessitate network-architecture and network-operations adaptations, so that the sensors' survivability limitations do not interfere with the required network overall performance, [2], [9].

In this paper, we study the deployment of Limited Sensing Stack Random Access Algorithms (LSSRAAs) in clustered wireless sensor networks containing high priority sensors, where all sensors have a limited life-span. The present study is an extension of that performed in [19]. We model sensor expiration, deploy LSSRAAs which give a delay advantage to high priority data and analyze system delays and sensor expiration rates. The organization of the paper is as follows: In Section 2, we present the system model. In Section 3, we outline the operations of the deployed LSSRAAs and summarize their performance in the absence of sensor expirations. In Section 4, we present numerical performance evaluations of the multi-cluster system in the presence of sensor expirations. In Section 5, we draw conclusions.

2 System Model

The overall system is depicted in Figure 1 and is comprised of multiple clusters connected with each other via a backbone network of Aggregate Forward Nodes (AFNs) and a Base Station (BS). Each cluster contains a large sensor population whose identities vary in time due to either mobility or expiration or both. Such a sensor population necessitates the deployment of a Random Access Algorithm (RAA) by the sensors for their data transmission within the cluster which contains them [23], [27]. The data generated across all clusters are in the form of identical length packets. Each sensor is a limited energy device whose stored energy may be exhausted by the retransmission and feedback sensing operations required for the transmission of a single packet (as dictated by the deployed RAA). In our model, we thus assume that each sensor generates a single packet and expires: if its stored energy is exhausted and the sensor expires before the successful transmission of the generated packet, the latter is lost/rejected; it is successfully transmitted, instead, if successful transmission precedes energy exhaustion and sensor expiration. The large sensor population, in conjunction with the model of a single

packet generated independently per sensor, gives rise to the *Limit Poisson User Model* per cluster [23], [27]. That is, the packet traffic per cluster is modeled as being generated by infinitely many independent Bernoulli users whose aggregate generated packet traffic is a Poisson process, where each packet is an independent user.

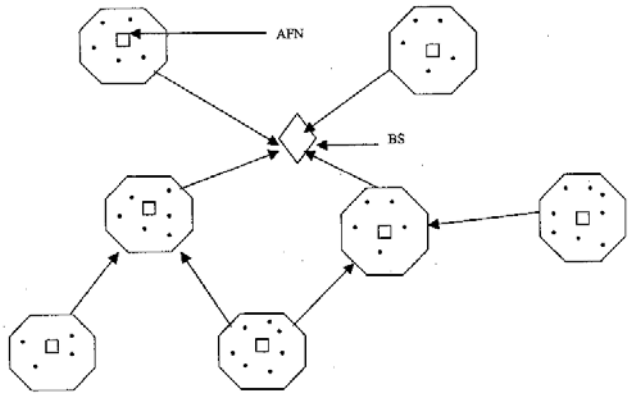


Figure 1: System Topology

It is assumed that each cluster in the network includes a distinct dedicated transmission channel (distinct dedicated frequency band) and contains two classes of sensors: *regular* and *high priority.* The regular sensors produce regular data and are not located close to the cluster boundaries. The high priority sensors are either located close to the cluster boundaries and may move shortly to the transmission range of another cluster, or they generate high priority data. The regular sensors monitor and transmit through only the transmission channel of the cluster they are contained in, while the high priority sensors may in addition monitor and transmit through the transmission channels of neighboring clusters. As we will see in Section III, the purpose of this additional channel monitoring employed by the high priority sensors is delay reduction, while their energy consumption may be simultaneously accelerated.

Each cluster transmission channel accommodates synchronous transmissions, where the time of all channels across the various clusters is slotted synchronously to identical length slots, each corresponding to a single packet transmission. Time is measured in slot units, where slot t occupies the time interval (t-1, t]. Slot feedback outcomes are assumed binary *Collision (C)* vs. *Non-Collision (NC)* across all cluster transmission channels: that is, it is assumed that a sensor monitoring a transmission channel may distinguish accurately between collision versus non-collision per channel slot, where collision indicates the simultaneous transmission by at least two packets, while non-collision indicates that a slot is either unoccupied or it is occupied with a single packet transmission. In addition, a collision results in complete loss of all involved packets requiring subsequent retransmissions, while a NC slot occupied with a single packet implies noiseless transmission. Given the transmission channel of a cluster, we will denote by x_t the feedback outcome (*C* vs. NC) of channel slot t.

Sensors transmit via given cluster channels deploying a stable Random Access Algorithm (RAA). As we will discuss in detail in Section III, the operations of such an RAA impose *initialization* periods during

which a data generating sensor only monitors passively the feedback outcomes of the channels, until the sensor enters a Collision Resolution Interval (CRI) within a single channel. During this CRI, the sensor both monitors feedback outcomes and occasionally retransmits, until the time when a successful retransmission occurs. As we will discuss in detail in Section III, for regular sensors, the initialization period and the CRI both involve only one channel: the transmission channel of the cluster that contains them. For high priority sensors, on the other hand, M > 1 channels are involved in the initialization period; the transmission channel of the sensor's home cluster plus the transmission channels of M-1 neighboring clusters, while the CRI still involves a single transmission channel selected by the sensor among the above M channels via some selection process (explained in Section III).

As in [19], we adopt a linear model for the energy consumption by each sensor in the system, as expressed by inequality (1) below, where we first define:

β: *The amount of per sensor energy consumed by the monitoring of a single channel slot.*

η: *The amount of per sensor energy consumed by a single packet (re) transmission.*

τ_1: *The number of slots during which a sensor passively monitors M channels in the system without attempting transmissions (M > 1 for high priority sensors).*

τ_2: *The number of slots during which a sensor monitors a single channel in the system, while participating in a Collision Resolution Interval (CDI).*

m: *The number of transmissions during the CRI the sensor participates in.*

ξ: *The total energy stored in a single sensor for packet transmission.*

Using the above notation, we assume that the sensor expires, with subsequent loss/rejection of its generated packet, If:

$$\beta\,(M\tau_1 + \tau_2) + \eta m \; > \xi \qquad (1)$$

We note that in the above expression we have assumed that the energies that a high priority sensor consumes for either monitoring or transmitting through a neighboring transmission channel are the same with those corresponding to its local transmission channel. This assumption may be untrue is some cases, where considerable variations in communication range may cause increased energy consumption when neighboring transmission channels are considered. In the latter cases, the quantities β and η defined above will be channel dependent, instead.

3 The Deployed LSSRAAs

As discussed in Section 2, the sensor population per cluster is time varying due to mobility and expirations. Thus, the sensor identities are not known to the system at all times. This gives rise to the unknown user population model which necessitates the deployment of Random Access Algorithms (RAAs) for transmission. At the same time, the per cluster Limit Poisson User Model adopted in Section II excludes the deployment of ALOHA-based RAAs, since the latter induce throughput zero then, [28], [29]. We propose, instead, the deployment of the Limited Sensing Stack Random Access Algorithms (LSSRAAs) in [23], in conjunction with the selection policy in [25] for the high priority sensors. Below, we describe the proposed LSSRAAs and the selection policy, for clusters containing both regular and high

priority sensors, where, in addition to their local transmission channel, the high priority sensors may also access a number of neighboring cluster transmission channels.

3.1 LSSRAA Operations for Regular and High Priority Sensors

Let us consider a total of M distinct clusters and their corresponding transmission channels, indexed by j = 1,…,M. Let us then denote by xt(j) ; j=1,…,M the feedback outcome of slot t in channel j. The regular sensors in cluster j, only monitor the feedback outcomes xt(j) and transmit only through channel j. The high priority sensors in cluster j monitor, instead, all channel feed backs xt(j); j=1,…,M and select dynamically one of the M channels for transmission. The LSSRAAs in [23], in conjunction with the selection policy in [25], induce an initialization and a collision resolution processes, where the initialization process is different for regular versus high priority sensors. We will first explain the collision resolution process, since it actually dictates the initialization process.

3.2 Collision Resolution Process

The collision resolution process induces a sequence of subsequent Collision Resolution Intervals (CRIs) per transmission channel whose characteristics are dictated by the operations of the deployed RAA. In the case of the LSSRAAs in [23] and the Limit Poisson User Model, a parameter Δ is utilized which corresponds to the size of the arrival interval resolved by each CRI (packet arrivals within such arrival interval transmitted successfully during the CRI), where the value of Δ is optimized for throughput maximization. The placement of the Δ-size window on the arrival access is determined asynchronously by the sensors, as will be explained below in this section, and is named the *examined interval* of the CRI.

The LSSRAAs in [23] are algorithms whose collision resolution process may be depicted by a stack containing a finite number of cells. Let us consider such an algorithm which may be described by a K-cell stack. Then, during some CRI on a given channel, each participating sensor follows the collision resolution rules by utilizing a counter whose values lie in the set of integers, [1,2,…K]. We denote by r_t the counter value of some participating sensor at slot t. The K different possible values of the counter place the user in one of the K cells of a K-cell imaginary stack. When its counter value is 1, the sensor transmits; it withholds at K-1 different stages otherwise. When a CRI begins, all sensors whose packet arrivals lie in the Δ-length examined interval of the CRI set their counters at 1; thus, they all transmit within the first slot of the CRI. If the examined interval contains at most one packet, the first slot of the CRI is a non-collision slot and the CRI lasts one slot. If the examined interval contains at least two packets, instead, the CRI starts with a collision which is resolved within the duration of the CRI (all packets involved in the initial collision are successfully transmitted during the CRI) via the following rules, where x_t denotes the feedback outcome of the channel's slot t:

The sensor transmits in slot t, if and only if $r_t = 1$. The packet of a sensor is successfully transmitted in slot t, if and only if $r_t = 1$ and x_t = NC.

The counter values transition in time during the CRI, as follows:

If x_{t-1} = NC and $r_{t-1} = j$; j=2,3,..,K, then $r_t = j-1$

If x_{t-1} = C and $r_{t-1} = j$; j=2,3,…K, then $r_t = j$

If x_{t-1} = C and $r_{t-1} = 1$, then, r_t =

1 ; w.p. 1/K 2 ; w.p. 1/K 3 ; w.p. 1/K

. . .

K ; w.p. 1/K

From the above rules, it can be seen that, on a given cluster transmission channel, a CRI which starts with a collision slot ends with K consecutive non-collision slots, an event which can not occur at any other instant during the CRI. Thus, the observation of K consecutive non-collision slots signals the certain end of a CRI to all users in the system; it either signifies the end of a CRI that started with a collision or the end of a sequence of K consecutive length-one CRIs. Therefore, a sensor which arrives in the cluster without any knowledge of the cluster-channel feedback history can synchronize with the local LSSRAA's algorithmic operations upon the observation of the first K-tuple of consecutive non-collision slots. This observation leads to the asynchronous placement of the size-Δ CRI examined intervals' on the arrival axis that describes the initialization process, for the regular and the high priority sensors.

3.3 Initialization Process

In general, if a CRI on a given cluster transmission channel ends with slot t, the examined interval of the next CRI on the channel is selected with its right most edge placed K-1 slots to the left of slot t and it contains those packets whose local *updates* (updates corresponding to the transmission channel of the given cluster) fall in the interval (t- K+1-Δ, t-K+1). The generation of the updates is explained below

For *regular sensors* in a given cluster, the *updates* $\{t^k\}$ of a packet (one packet corresponds to one sensor) are generated strictly for the local transmission channel (that corresponding to the cluster where the sensors reside) and are as follows: Let t_0 be the slot within which a packet is generated. Then, define t^0 to be equal to t_0. Starting with slot t^0, the corresponding regular sensor senses continuously the feedback outcomes of the local transmission channel. It does so passively, until it observes the first K-tuple of consecutive NC slots, ending with slot t_1. If $t^0 \in (t_1 – K+1-\Delta, t_1 – K+1)$, the sensor participates in the CRI which starts with slot $t_1 + 1$ on the local channel. Otherwise, it updates its arrival instant to $t^1 = t^0 + \Delta$ and observes local feedback outcomes passively until the end of the latter CRI, ending with slot t_2. If $t^1 \in (t_2 – K+1-\Delta, t_2 – K+ 1)$, the sensor participates in the CRI which starts with slot t_2; otherwise, the sensor updates its arrival instant by Δ again and repeats the above process. In general, if $\{t_n\}$ $n \geq 1$ denotes the sequence of consecutive local CRI endings since the first K-tuple of consecutive NC slots, the sensor participates in the k^{th} CRI if $t^{k-1} \in (t_k – K+1-\Delta, t_k – K+1)$ and $t^n \notin (t_{n+1} – K+ 1 - \Delta, t_n – K+1)$; for all $n \leq k-2$. In the latter case, t_k denotes the beginning slot of the first after the packet's arrival CRI whose examined interval the packet's update falls into.

For *high priority* sensors in any of the M clusters, the initialization process develops as follows. Upon generating a packet, each high priority sensor starts observing continuously the feedback outcomes from all M transmission channels in the M clusters. Subsequently, the sensor generates M sequences of updates, separately for each channel, as described in the above paragraph, imagining itself as a regular user in each cluster. The high priority sensor participates then in the first across all M channels CRI whose examined interval its corresponding update falls into and transmits its packet successfully during the process of the latter. This dynamic CRI participation constitutes the high priority sensor's *selection policy.* When m updates (across m channels) of a high priority sensor fall within the examined intervals of m simultaneously starting CRIs (across m channels), one of the m CRIs is selected equiprobably.

3.4 Performance Characteristics in the Absence of Sensor Expirations

In the absence of sensor expirations and of high priority sensors in M clusters, the LSSRAAs in this section give rise to M identical and independent RAA systems. As found in [23], the throughput of each such system, for the Limit Poisson User Model, is then 0.43 (in average number of packets per slot) for all K values in the stack, where throughput is defined as the highest traffic Poisson rate for which the system is stable. For different K values, the 0.43 throughput is attained, however, for different values of the window size Δ. In [23], it was found that for K = 2 the optimal window size is $\Delta^* = 2.33$, while for K=3 the optimal window size is $\Delta^* = 2.56$.

In the absence of sensor expirations and the presence of high priority sensors in M participating clusters, their dynamic selection policy causes coupling of the RAA operations across the M cluster transmission channels. The result is throughput reduction of the M channel system, at the gain of reduced delays for the high priority traffic. An outline of the throughput evaluation process for any K value LSSRAA, in the presence of the Limit Poisson User Model can be found in [19]. In Figure 2, we plot, as an example, the stability regions induced for the LSSRAA with K=3 and M=2 clusters, as found in [25]. The latter figure corresponds to Limit Poisson Models for the two regular traffics (local to clusters 1 and 2 with corresponding Poisson rates λ_1 and λ_2), as well as for the high priority traffic (whose Poisson rate is λ_3). We note that, for M = 2 and changing K values, the stability regions in the figure will remain unchanged, as long as the corresponding to the chosen K value optimal window size Δ is used. Regarding delays, they change when either one of the M or K values varies. For K fixed, as M increases, the per channel throughput decreases, at the gain of enhanced delay reductions for the high priority traffic [25]. Quantitative delay results will be presented in Section 4.

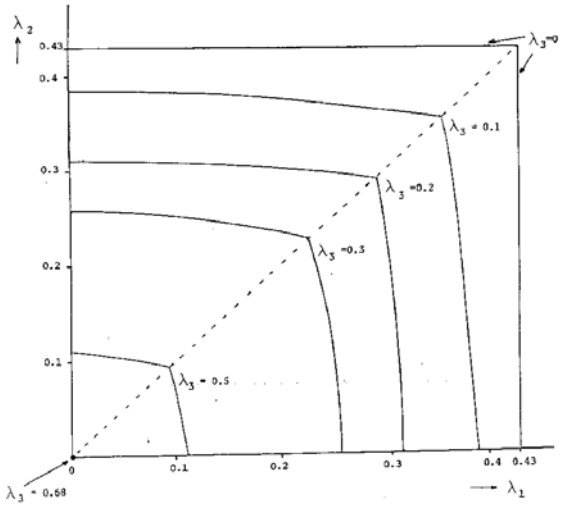


Figure 2: Stability Regions for K=2 and M=2

3.5 Performance and Dynamics in the Presence of Sensor Expirations

In the presence of sensor expirations, sensors may expire before the successful transmission of their generated packet, causing packet losses/rejections. System stability becomes then a void concept; thus, so does throughput. The relevant system performance criteria for both the regular and the high priority traffics are then: rejection rates (percentage of rejected traffic) and delays of the successfully

transmitted packets, all as functions of the input Limit Poisson traffic rates. In Section IV below, we include simulated such performance results, for Limit Poisson input traffics with varying rates and the sensor expiration model in (1) with various parameter values.

The deployed LSSRAAs, in conjunction with the power consumption formula for the sensors, induce specific packet rejection rates and specific delays of the successfully transmitted packets, as functions of the input traffic rates. As a result, required constraints on packet rejection rates and delays dictate then the acceptable operational rate regions of the input traffics. Subsequently, the latter regions determine the size of each cluster, where cluster size refers to the aggregate data rate generated in it rather than its geographical area**.** We will discuss specific such arising cluster size issues in Section 4.

4 Numerical Performance Evaluations

We simulated the case of seven clusters indexed from 1 to 7. We assumed that each cluster contains the same fraction α of high priority users and that each such user monitors the same number M-1 of neighboring transmission channels (in addition to that in its own cluster). Visualizing the placement of clusters 1 to 7 on a circle, a high priority user monitors n neighboring channels in each one of the clockwise and counterclockwise directions; if M-1= 2n, while it monitors n+1 in the clockwise and n in the counterclockwise directions neighboring channels; if M-1=2n+1. All traffics have been modeled as Poisson, where system performance in both the presence and the absence of sensor expirations has been evaluated. In the presence of sensor expirations, various values of the parametersβ, η and ξ in the expiration formula (1) have been considered, while, here, we present the results for a representative selection subset. For both the regular and the high priority traffics, we present expected delay results in all cases, while we also present rejection rates results, when the sensor expiration formula is active.

In both the presence and the absence of sensor expirations, we tested the following high priority traffic fraction values: α = 0.05, 0.1, 0.3, 0.5. We also then tested the following values for the number of channels monitored by high priority users: M=2, M=3, M=4, M=5. In the presence of sensor expirations, the selected sets [β, η, ξ] of constants' values in formula (1) were: [5, 30, 300], [5, 10, 100], [5, 50, 500] and [5, 50, 1000].

In the absence of sensor expirations, representative results are depicted by Figures 3 and 4. In the figures' legend, 7-k-R/S-L/M-NE, 7 depicts the 7 channel system, k=M-1, R stands for regular while S stands for high priority users, L/M depicts the fraction of high priority traffic (0.05 / 0.3) and NE stands for absence of energy constraints. In Figure 3, expected delays for both the regular and the high priority users are plotted against cluster traffic rates, when the fraction α of high priority traffic per cluster is 0.05. In Figure 4, the expected delay results are for α = 0.3, instead. Our results exhibit the significant delay advantage of the high priority traffic, as compared to that of the regular traffic, where this advantage increases as the traffic rate does. In addition, as the fraction of high priority traffic increases, the expected delays of the latter traffic decrease, while the effect of increasing number M of monitored channels simultaneously decreases; in all cases, the latter effect is non-negligible only for relatively high traffic rates. For example, for α = 0.3, it suffices to assign a single neighboring channel to the high priority traffic (M-1=1), for monitoring and possible transmission, where then the expected delays of the latter traffic never exceed 6.5 slots.

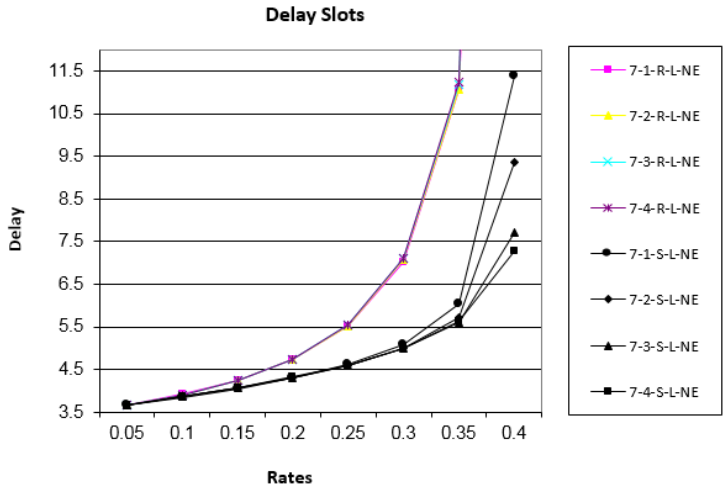


Figure 3: Expected delays in the absence of energy constraints as functions of cluster traffic rates, for α = 0.05

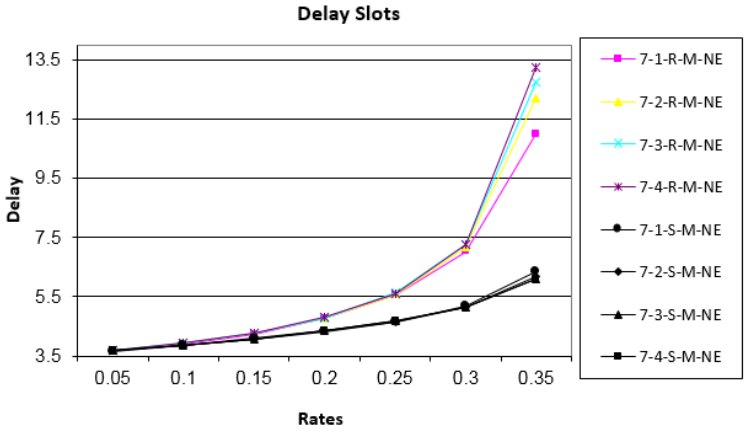


Figure 4: Expected delays in the absence of energy constraints as functions of cluster traffic rates, for α = 0.3

In the presence of sensor expirations, representative results are depicted by Figures 5 to 12. The figures' legend is as that of Figures 3 and 4, where E (instead of NE) stands here for the presence of energy constraints. Figures 5, 7, 9 and 11 display expected delays as functions of cluster traffic rates, while Figures 6, 8, 10 and 12 display rejection rates, instead. Figures 5, 6, 7 and 8 show results for α = 0.05, while Figures 9, 10, 11 and 12 exhibit results for α = 0.3. Figures 5, 6, 9 and 10 correspond to expiration constants' values [β, η, ξ] = [5, 10, 100], while Figures 7, 8, 11 and 12 correspond to [β, η, ξ] = [5, 50, 500], instead. From the results in the figures, we first observe that, in all cases, as the traffic cluster rate increases, expected delays stabilize when rejection rates are sufficiently large; for the case of [β, η, ξ] = [5, 50, 500], where the stored per sensor energy is relatively high, the expected delays for the regular users take a dip at the point where the expected delays assume a sharp increase. As expected, in all cases, the rejection rates for the high priority users increase, as the number M of monitored channels increases, since these users loose then increased energy in monitoring. As the stored per sensor energy increases (from ξ = 100 to ξ = 500), the expected delays increase; since less traffic is then rejected. At the same time, when the stored per sensor energy is relatively large, increase in the number M of monitored channels does not present the high priority users with a significant delay advantage, while it reduces significantly their rejection losses in the presence of relatively high cluster

traffic rates and relatively low α values; when M is larger, the monitoring time by the high priority users is then sufficiently low to allow transmission before expiration. As compared to the regular users with relatively high energy reserves, high priority users possessing the same reserves are presented with a significant delay advantage and with simultaneously minimal losses; for cluster traffic rates below 0.4, while they experience less significant delay and loses reductions; for cluster traffic rates above 0.4. The tradeoff between delays and rejection rates is strongly present when the stored per sensor energy is low, where, as the number M of monitored by the high priority users increases, delays for all users decrease and rejection rates increase, while the delay and rejection differences between regular and high priority users also increase; as M increases, high priority traffic experiences increased delay advantages at the expense of also increases rejection losses.

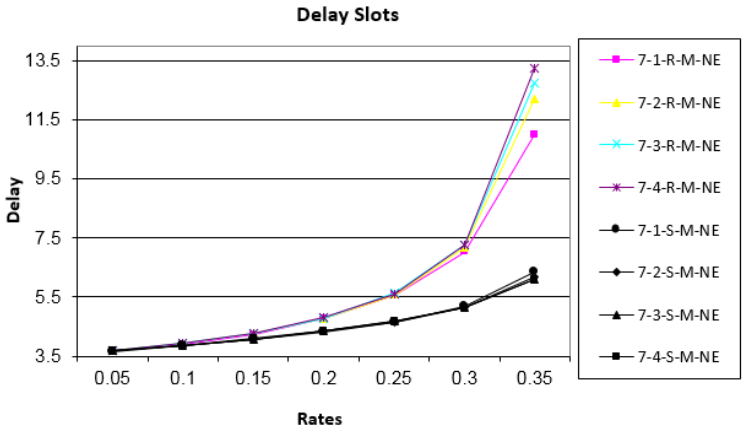


Figure 5: Expected delays in the presence of energy constraints as functions of cluster traffic rates, for α = 0.05 and [β, η, ξ] = [5, 10, 100]

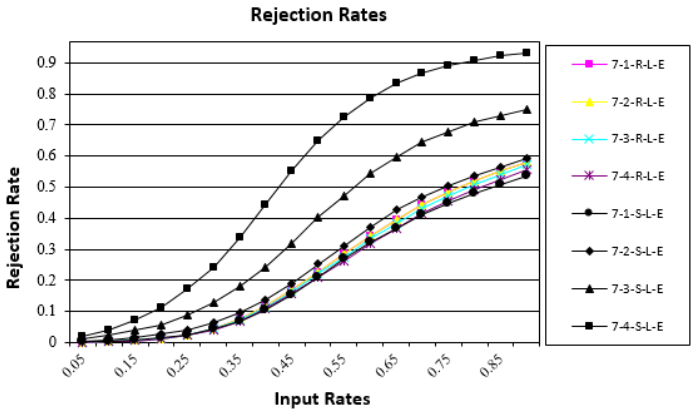


Figure 6: Rejection rates in the presence of energy constraints as functions of cluster traffic rates, for α = 0.05 and [β, η, ξ] = [5, 10, 100]

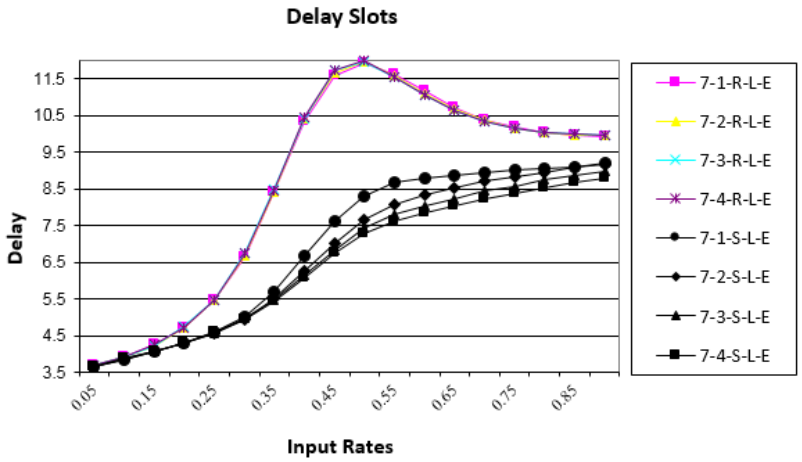


Figure 7: Expected delays in the presence of energy constraints as functions of cluster traffic rates, for $\alpha = 0.05$ and $[\beta, \eta, \xi] = [5, 50, 500]$

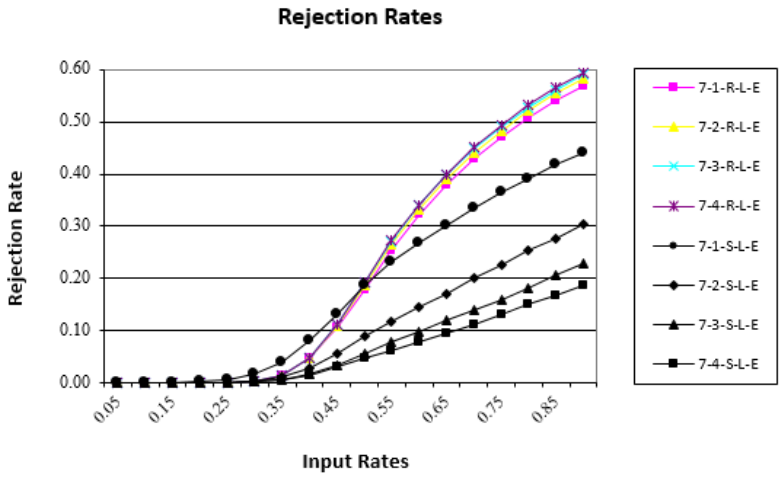


Figure 8: Rejection rates in the presence of energy constraints as functions of cluster traffic rates, for $\alpha = 0.05$ and $[\beta, \eta, \xi] = [5, 50, 500]$

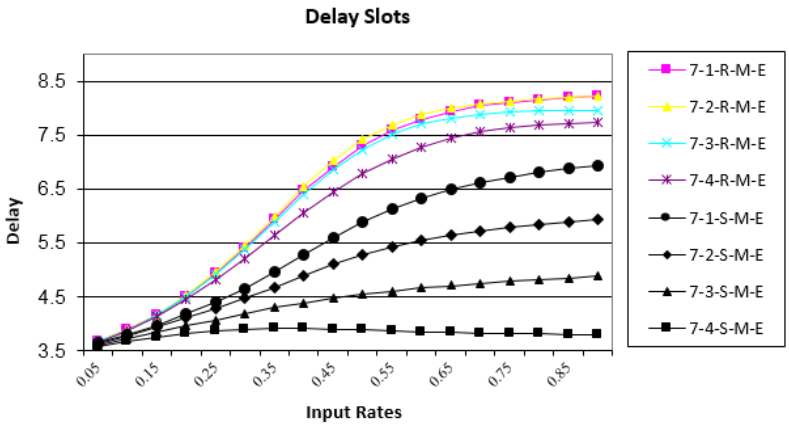


Figure 9: Expected delays in the presence of energy constraints as functions of cluster traffic rates, for $\alpha = 0.3$ and $[\beta, \eta, \xi] = [5, 10, 100]$

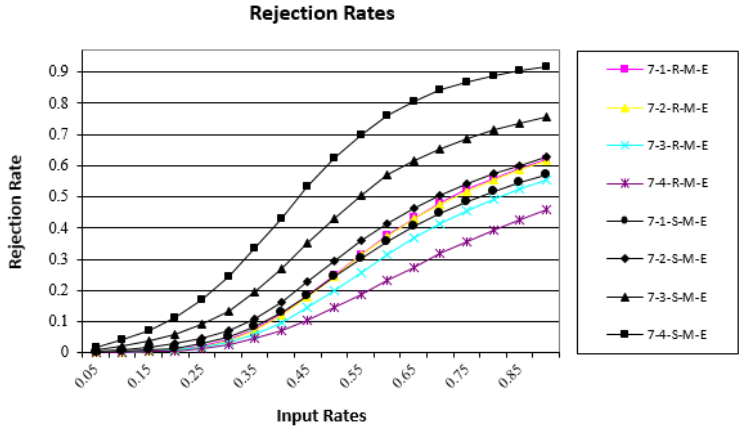


Figure 10: Rejection rates in the presence of energy constraints as functions of cluster traffic rates, for $\alpha = 0.3$ and $[\beta, \eta, \xi] = [5, 10, 100]$

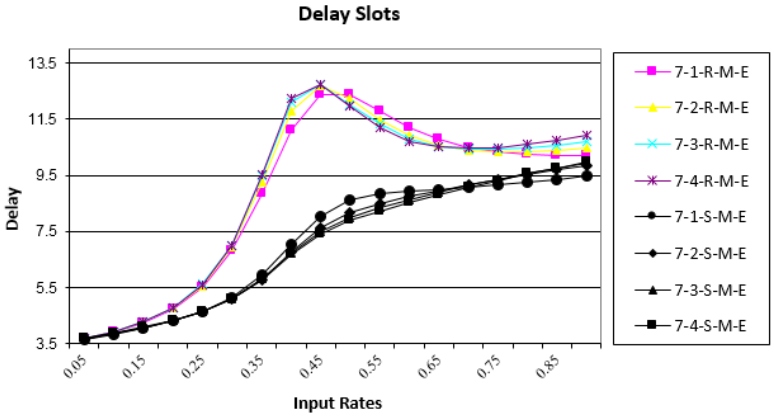


Figure 11: Expected delays in the presence of energy constraints as functions of cluster traffic rates, for $\alpha = 0.3$ and $[\beta, \eta, \xi] = [5, 50, 500]$

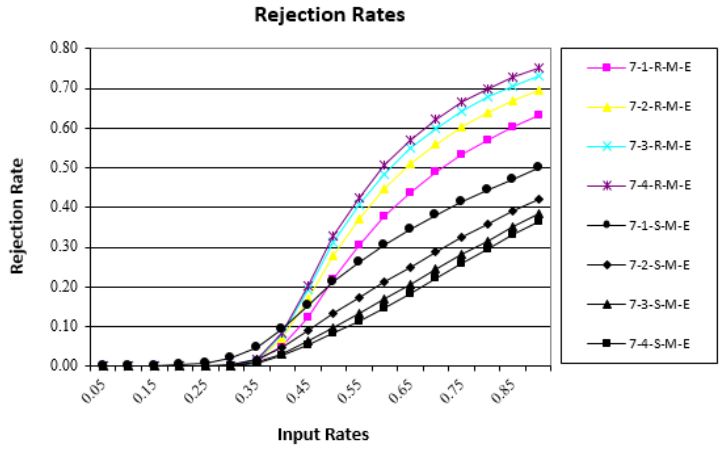


Figure 12: Rejection rates in the presence of energy constraints as functions of cluster traffic rates, for $\alpha = 0.3$ and $[\beta, \eta, \xi] = [5, 50, 500]$

5 Conclusions

We analyzed and numerically evaluated a random access algorithm for deployment in clustered sensor networks which incorporate regular as well as high priority traffics, where high priority users monitor and occasionally transmit through neighboring cluster channels for delay reduction All sensors are assumed to contain limited energy reserves with energy consumed by channel monitoring and transmission attempts; such limited reserves inevitably induce traffic rejections. Our results show significant difference in system behavior between low level versus relatively high level of stored energy per sensor: (a) In the case of low level of energy reserves, a profound tradeoff between delays and traffic rejection rates presents itself. As the number of channels monitored by the high priority users increases, delays for all users decrease and rejection rates increase, where the delay and rejection rate differences between regular and high priority also increase: high priority users experience then lower delays, at the expense of increased rejection rates. (b) In the case of relatively high energy reserves, the system behavior may be distinguished between two regions of cluster traffic rates. For cluster traffic rates less than 0.4, the delays of the high priority users are significantly lower than those of the regular users, while all users simultaneously experience low traffic rejection rates; increase in the number M of the monitored by the high priority users channels presents then no advantage, M-1 = 1 being recommended. For cluster traffic rates above 0.4, the delays of the high priority users are not significantly lower than those of the regular users, while the rejection rates of the high priority users are then significantly lower than those of the regular users; the latter rejection rate reduction increases with increasing number M of monitored by the high priority users channels and decreasing fraction of high priority traffic.

REFERENCES

[1]. D. Blough and S. Paolo, “Investigating Upper Bounds on Network Life-Time Extension for Cell-Based Energy Conservation Techniques in Stationary Ad Hoc Networks”, in Proc*. ACM MobiCom,* Atlanta, GA, Sep. 23-28, 2002, pp. 183-192.

[2]. M. Bhardwaj and A.P. Chandrakasan, “Bounding the Lifetime of Sensor Networks via Optimal Role Assignments”, in Proc*. IEEE INFOCOM,* New York, Jun. 23-27, 2002, pp. 1587-1596.

[3]. J.H. Chang and L. Tassiulas, “Energy Conserving Routing in Wireless Ad Hoc Networks”, in Proceedings of *IEEE INFOCOM*, Tel Aviv, Israel, Mar. 26-30, 2000, pp. 22-31.

[4]. Y. T. Hou, Y. Shi and H.D. Sherali, “On Node Lifetime Problem for Energy-Constrained Wireless Sensor Networks”, *ACM/Springer Mobile Netw. Applicat.,* vol. 10, no. 6, pp. 865-878, Dec. 2005.

[5]. Y.T. Hou, Y. Shi, J.H. Reed and K. Sohraby, “Flow Routing for Variable Bit Rate Source Nodes in Energy-Constrained Wireless Sensor Networks”, in Proc. *IEEE International Conference on Communications,* Seoul Korea, May 16-20, 2005, pp. 3057-3062.

[6]. V. Rodoplu and T.H. Meng, “Minimum Energy Mobile Wireless Networks”, *IEEE J. Sel. Areas Commun.,* vol. 17, no. 8, Aug. 1999, pp. 1333-1344.

[7]. K. Sohrabi, J. Gao, V. Ailawadhi and G. Pottie, "Protocols for Self-Organizing of a Wireless Sensor Netrwork", *IEEE Pers. Communications Magazine,* vol. 7, Oct. 2000, pp. 16-27.

[8]. V. Srinivasan, P. Nuggehalli, C.F. Chiasserini and R. Rao, "Cooperation in Wireless Ad Hoc Networks", in Proc. *IEEE INFOCOM,* San Francisco, CA, Mar. 30-Apr. 3 2003, pp. 808-817.

[9]. R. Wattenhofer, L. Li, P. Bahl and Y.M. Wang, "Distributed Topology Control for Power Efficient Operation in Multihop Wireless Ad Hoc Networks", in Proc. *IEEE INFOCOM,* Snchorage, AK, Apr. 22-26, 2001, pp. 1388-1397.

[10]. A. T. Burrell and P. Papantoni Kazakos, "Performance Monitoring in Sensor Networks". *ITNG 2009*, April 27-29, 2009, Las Vegas, Nevada.

[11]. Y. T. Hou, Y. Shi and H.D. Sherali, "Rate Allocation and Network Lifetime Problems for Wireless Sensor Networks", *IEEE Trans. Networking,* vol. 16, no. 2, April 2008, pp. 321-334.

[12]. D. Bertsekas and R. Gallager, *Data Networks*, Englewood Cliffs, NJ, Prentice Hall, 1992.

[13]. H. Luss and D.R. Smith, "Resource Allocation among Competing Activities: A Lexicigraphic Minimax Approach". *Operations Research Letters,* vol. 5, no. 5, Nov. 1986, pp. 227-231.

[14]. P. Papantoni Kazakos and A.T. Burrell "The Implementation of Dynamic Rate Allocation in Sensor Networks" *Journal of Intelligent and Robotic Systems,* 2010, Vol. 58, pp. 211-238.

[15]. A. T. Burrell and P. Papantoni-Kazakos, "Extended Sequential Algorithms for Detecting Changes in Acting Stochastic Processes", *IEEE Trans. System, Man and Cybernetics,* Vol. 28, no. 5, Sept. 1998, pp. 703-710.

[16]. A. T. Burrell, D. Makrakis, and P. Papantoni-Kazakos, "Traffic Monitoring for Capacity Allocation of Multimedia Traffic in ATM Broadband Networks", *Telecommunication Systems*, Vol. 9, July 1998, pp. 173-206.

[17]. A.T. Burrell and P. Papantoni-Kazakos, "On-Line Learning and Dynamic Capacity Allocation in the Traffic Management of Integrated Services Networks", *European Transactions on Telecommunications,* Special Issue on Architectures, Protocols, and Quality of Service for the Internet of the Future, Vol. 10, No. 5, March/April 1999, pp. 202-214.

[18]. Fatma Salem, A.T. Burrell and P. Papantoni-Kazakos, "Dynamic Architectural Reconfigurations of Sensor Networks", *IEEE IECON'2010*, Glendale AZ, Nov. 7-10, 2010.

[19]. Fatma Salem, A.T. Burrell and P. Papantoni-Kazakos, "Dynamic Architecture- Reconfiguration Algorithms and Transmission Protocols for Clustered Sensor Network Topologies with Prioritized Data," *International Scholarly Research Network (ISRN) Sensor Networks,* 2012, Article ID 452981, 17 pages, doi: 10.5402/2012/452981.

[20]. S. Bandyopadhyay and E.J. Coyle, "An Energy Efficient Hierarchical Clustering Algorithm for Wireless Sensor Networks", *IEEE INFOCOM 2003.*

[21]. D. Kazakos and P. Papantoni-Kazakos, *Detection and Estimation,* Computer Science Press, New York, NY, 1990.

[22]. A. Papoulis and S.U. Pillai, *Random Variables and Stochastic Processes,* 4th Edition, McGraw-Hill, New York, NY, 2002.

[23]. A.T.Burrell and T. Papantoni, "A Class of Limited Sensing Random Access Algorithms with Resistance to Feedback Errors and Effective Delay Control", *Journal of Communications and Networks,* 2006, Vol. 8, No. 1.

[24]. P. Papantoni-Kazakos, "Multiple Access Algorithms for a System with Mixed Traffic: High and Low Priority," *IEEE Trans. Comm*., March 1992, Vol. 40, pp. 541-555.

[25]. P. Papantoni-Kazakos, H. Delic, M. Paterakis, and M. Liu, "Transmission Algorithms for a Multi-Channel Packet Radio System with Priority Users," *International Journal of Digital and Analog Communication Systems (IJDACS) John Wiley & Sons*, October-December 1993, Vol. 6, No. 4, pp. 193-212.

[26]. M. Paterakis and P. Papantoni-Kazakos, "A Simple Window Random Access Algorithm with Advantageous Properties," *IEEE Trans. Inform. Th.,* September 1989, Vol. 35, pp. 1124-1130.

[27]. I. F. Akyildiz, W.Su, Y. Sankkarasubramaniam and E Cayirci, "Wireless Sensor Networks: A Survey", *Computer Networks (Elsevier),* vol. 38, no. 4, pp. 393-422, 2002.

[28]. A. T. Burrell and P. Papantoni-Kazakos, "Wireless Networks in Hybrid Topologies: Signaling, Transmission, and Dynamic Capacity Allocation Viewed in an Integrated Fashion," *ITC Mini-Seminar Performance Modeling and Design of Wireless/PCS Networks*, Cambridge, MA, October 3, 1996.

[29]. A.T. Burrell and P. Papantoni-Kazakos, "Random Access Algorithms in Packet Networks-A Review of Three Research Decades", *International Journal of Communications Network and System Sciences (IJCNS),* October 2012, Vol. 5, No. 10..

[30]. S. P. Meyn and R.L. Tweedie, *Markov Chains and Stochastic Stability,* Springer-Verlag, 1993.

[31]. M. Kaplan, "A Sufficient Condition for Nonergodicity of a Markov Chain", *IEEE Transactions on Information Theory*, vol. 25, 1979, pp. 470-471.

TNC Transactions on Networks and Communications
SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
VOLUME 2, ISSUE 5
ISSN: 2054 - 7420

Modeling Structural Behaviour of Inhibitors of Cloud Computing: A TISM Approach

[1]Ambikadevi Amma. T, [2]N. Radhika and [3]Pramod V.R.
[1]Karpagam University, Coimbatore, India
[2]Computer Science & Engg Dept, Amrita University, Coimbatore, Tamil Nadu, India
[3]Dept. of Mechanical Engineering, NSS College of Engg, Palakkad, Kerala, India
prof.ambikadevit@gmail.com; n_radhika@cb.amrita.edu; pramodram09@gmail.com

ABSTRACT

Cloud computing is the delivery of computer resources over a network through web browsers, while the actual physical location and organization of the equipment hosting these resources are hidden from the users. Some of the IT organizations are undergoing severe budgetary constraints depends on clouds for the infrastructure and services. The major attributes of cloud computing are multi-tenancy, massive scalability, elasticity, pay as you use and self-provisioning of resources of the cloud. Cloud computing strategy is subjected to many inhibitors. For finding the interrelationship among inhibitors ISM (interpretive structural modeling) is used which a well is proved technology for finding the interrelationship among elements. An innovative version of interpretive structural model is known as Total Interpretive Structural Model (TISM). In Total Interpretive Structural Modeling (TISM), influence/enhancement of inhibitors and their interrelationship is considered. Total interpretive structural model consists of the following steps. They are identification of elements, pair-wise comparison, level partition, interaction formation, diagraph representation and diagrammatic representation of total interpretive structural model. The methodology of TISM is used to delineate the hierarchical relationship of inhibitors of cloud computing.

Keywords: Cloud computing, Inhibitors, Partition levels, Interaction matrix, TISM.

1 Introduction

Cloud computing is enhancing the effectiveness of computer services while operating in the furious industrial environment. Cloud service providers comply with strict operation policies and measurements to minimize failures in the system. The strategies of cloud computing is subjected to many inhibitors. Cloud computing facility is availed by the customers through internet by using any web browsers. Cloud computing services are having many issues. As these issues are broken with innovation, cloud will move from consumers of small medium business to larger and larger enterprise deployments. Top challenges of cloud computing are security, performance, availability and integrity of data. A great deal of uncertainty is pertained about the security at different levels of network. Network security solutions are not in tolerance with the movement required for cloud to deliver its promise and cost efficiencies. A large number of servers may be present in the cloud .The potential to consolidate millions of servers into dynamic meshes is the biggest pay-off cloud computing. Market valuation and growth potential

DOI: 10.14738/tnc.25.390
Publication Date: 23th September 2014
URL: http://dx.doi.org/10.14738/tnc.25.390

depend on the security, physical infrastructure and network management. Privacy is the accountability to data and transparency to an organization practice about personal information existing. Compliance requirements impact in many ways. Cloud can have cross multiple jurisdictions .Data may be stored in different countries or may be in different states.

Cloud Computing is a new scenario in which customers can use the services and infrastructures by paying an amount for their usage. This is beneficial to some of the IT organizations undergoing severe budgetary constraints for the development of infrastructures and enhancement of hardware and software. The core technologies used in cloud are web applications, services, virtualization and cryptography. The services rendered through cloud are Software as a Service (SaaS), Infrastructure as a Service (IaaS), and Platform as a Service (PaaS). In Software as a Service, the service providers provide the users the service of using any type of application software. In IaaS, the service providers provide the networking equipment, storage backups and servers. In PaaS the service providers provides the platform to the users any type of operating systems along with hardware and Software.

Cloud computing is the delivery of computing resources as a service rather than as a product, whereby information is provided to computers and other devices as a utility over a network. Cloud computing describes a new supplement , consumption, and delivery model for IT services based on internet protocols, and it typically involves provisioning of dynamically scalable and often virtualized resources. Cloud computing providers deliver application via the internet, which are accessed from a web browser.

2 Literature Survey

Cloud computing is becoming a well-known buzz word now a days. Privacy issues and security problems are pointed out as barriers for users to adopt into cloud computing systems. Users of cloud computing worry about their business information and critical IT resources in the cloud computing systems which are vulnerable to be attached [1]. Cloud computing allows providers to develop, deploy and run applications that can easily grow in capacity work rapidly without any concern on the properties and the locations of the undergoing infrastructures [2]. Availability is one of the goals of security. It ensures the user to use them at any time at any place. Hardening and redundancy will enhance the availability of the cloud system. Many cloud computing system provide cloud infrastructure and platforms based on virtual machine [Farzad Sabahi,2011] Confidentiality means keeping user's data secret in the cloud system.

Data integrity means prurience the information. No change or no modification by unauthorized users. Access control is another good in security. Access control means to regulate the use of the system including applications, infrastructure and data. Auditing is another phenomenon that could be added as an additional layer above the virtualized OS hosted on the virtual machine [7].Secret information of individual users and business are stored and managed by the service providers.

Cloud computing raises a range of important policy issues, which include issues of Privacy, security, anonymity, telecommunications capacity, government surveillance, reliability, and liability, among others. Users will expect Reliability and Liability on the cloud service resource. Especially if a cloud provider takes over the task of running "mission-critical" applications and will expect clear delineation of liability if serious problems occur. Users will expect that the cloud provider will prevent unauthorized access to both data and code, and sensitive data will remain private. Users will expect to be able to access and use the cloud where and when they wish without hindrance from the cloud provider or third

parties, while their intellectual property rights are upheld. Each of these interrelated issues will be considered in terms of its importance, what realistic expectations users might have, and the policy implications.

S. Pearson et al describe privacy manager mechanism in which user's data is safe on cloud. In this technique the user's data is in encrypted form. Privacy manager make readable data from result of evaluation manager. In obfuscation data is not present on Service provider's machine so there is no risk with data, so data is safe on cloud but this solution is not suitable for all cloud applications. When input data is high these methods require a large amount of memory [2]. In [3], the authors present procedural and technical solution; both are producing solution to accountability to solve security risk in cloud. Here policies are decided by the parties that use, store or share that data irrespective of the jurisdiction in which information is processed. But it has limitation that data processed on SP is in unencrypted form at the point of processing .So there is a risk of data leakage. In [4], the author gives a language which permits to serve data with policies by agent; agent should provide their action and authorization to use particular data. In this logic data owner attach policies with data, which contain a description of which actions are allowed with each data. In [5], authors give a three layer architecture which protects information leakage from cloud. It provides three layers to protect data. In first layer the service provider should not view confidential data, in second layer service provider should not do the indexing of data, in third layer user specify use of his data and indexing policies. In [6], authors present accountability in federated system to achieve trust management. The trust towards use of resources is accomplished through accountability so to resolve problem for trust management. In federated system they have given three layer architecture, First layer is authentication and authorization. Public key cryptography is used in first layer. Second layer is accountability which perform monitoring and logging. The third layer is anomaly detection which detects misuse of resources. This mechanism requires third party services to observe network resources.

Interpretive structural modeling is a well-established methodology for identifying relationships among elements present in a complex structure [15]. ISM is an interactive learning process in which set of directly or indirectly related elements are structured into a comprehensive model. For identity relationships among items, the ism methodology can be established. The variables in the specific problem or issue are identified first and then a contextually relevant subordinate relation is taken. Based on pair-wise comparison of variables, a structural self-interaction matrix (SSIM) is developed from the element set. Transitivity is checked and a matrix model is obtained. ISM is derived from the partitioning of the element and an extractive of the structural model [10]. In this approach conceptual and computational leverage are exploited to explain the contextual relationship among a set of variables. According to Warfield [11] a set of requirement are needed for interpretive structural modeling. They are: a) Inclusion of scientific elements b) A complex set of relation can be exhibited c) Complex set of relations permits continuous observation, questioning and modification d) Consequence with perceptions and analytical process of the originators e) Public audience can early learn.

3 Methodology

ISM is an interpretive learning process. Judgment of the group decides the relationship of different elements in the system. Over all structure is extracted from the set of elements on the basis of mutual relationship hence it is structural and overall structure is portrayed in a diagraph model hence it is a

modeling technique. Total Interpretive Structural Model has some common steps of ISM. Reachability and partition levels are adopted as it is in the process of interpretive structural model. Steps for obtaining Total Interpretive Structural Model is briefly described below.

3.1 Step I: Inhibitors identification and its definition

First step in any structural modeling is identification and definition of elements whose relationship is to be studied. For these purpose inhibitors of cloud computing is identified from literature survey and discussions with domain experts and is shown in Table1.

Table 1: Inhibitors and definitions.

In. No.	Name of inhibitors	Definition
1	lack of sufficient security	Security in cloud has boarder area comprising software security hardware security
2	Lack of reliability	Data taken from clouds should be correct in all sense. Customer can rely on the service of cloud
3	Lack of portability	It is important to map out the dataflow from the current infrastructure to an eventually cloud service provider whether private or public cloud.
4	Lack of privacy	Privacy is the accountability to data and transparency to an organization. Personal information should be kept constant.
5	Lack of standardization	Standards are scares with in the cloud. Mass adoption is difficult without standards.
6	Lack of comprehensive management tool	These tools would help automatic service provisioning balance workloads and aids with capacity planning and configuration management
7	Week access control	We will have more control to manage and racking of servers, networking and cabling as well as security
8	Lack of data confidentiality	Customers should have fading that every service getting through cloud is correct and secure. Hence confidence on cloud will increase
9	Ineffective backup management	Since everything depend on computation any error can cause defects or destruction .Hence a backup is very necessary and new not available
10	Cost/time barrier	It should be evaluated closely as a cloud migration could actually be much more feasible realistic and less expensive then companies actually realize
11	.Network management barrier	Changes in the security requirement may change the topology of network.
12	Legal issues.	Cloud services may be among countries. .Different countries are having different jurisdiction which may affect cloud services.
13	Infrastructure security at network level	For using public cloud the topology of the network may vary with the security requirement
14	Infrastructure security at host level	Virtualization security threads like VM escape, system configuration drifts and inside threads
15	Infrastructure security at application level	Application security spectrum should be considered web browser security should be taken into account
16	Lack of data integrity	Data should be correct in all means so that integrity validation should be applied
17	Lack of data availability	Customer requirement may be satisfied by making the data in cloud availability to everyone
18	Lack of data security	Data security includes the security of stored data and retrieving data.

3.2 Step II: Contextual relationship definition

Structural Self Interaction matrix is developed by relating elements contextual relationship. Contextual relationship between different inhibitors is studied. Inhibitor 1 is compared with all other inhibitors and study .how inhibitor1 influence/enhance inhibitor 2 and how inhibitors 1 influence/enhance inhibitor 3 etc. To capture the contextual relationship among inhibitors experts opinion is solicited.

3.3 Step III: Interpretation of relationship

In traditional ISM contextual relationship remains silent. Relationship alone is charted out but not given importance on how that relationship really works. In TISM explanation of how the inhibitors influence /enhance with each other is considered. It also explain in what way they influence /enhance each other.

3.4 Step IV: Pair-wise comparison

A pair-wise comparison of elements is used to develop SSIM (Structured Self interaction Matrix.) In formal ISM interpretation indicate direction of relationship only when there is relation among elements. When there is relation from i to j -V, j to i -A, i to j and j to i -X and for no relation O is used. TISM make use of the concept by answering the interpretive query in step III. For each paired comparison, first element should be compared with all the remaining elements .For each comparison the entry should be Y for relation or N for no relation. The reason for Y should be provided. Comparing all the row elements, a paired relationship in the form of interpretive logic –knowledge base is obtained and is shown in Table 2.

Table 2: Interpretive logic –knowledge base

Inhibitor No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
I	Y	Y	N	N	N	N	Y	Y	N	N	Y	N	Y	Y	N	Y	Y	Y
2	N	Y	N	N	Y	N	N	Y	N	N	N	Y	N	N	N	Y	Y	Y
3	N	N	Y	N	N	N	N	N	N	N	N	Y	N	N	N	N	Y	Y
4	Y	Y	N	Y	N	N	Y	Y	N	N	N	Y	Y	N	N	Y	Y	Y
5	N	N	N	N	Y	Y	N	Y	N	N	N	Y	N	N	N	N	Y	Y
6	N	Y	N	N	Y	Y	N	Y	N	N	N	Y	N	N	N	Y	Y	Y
7	N	N	Y	Y	Y	Y	Y	Y	N	N	N	Y	Y	Y	Y	Y	Y	Y
8	Y	N	N	Y	Y	N	Y	Y	N	N	N	Y	Y	Y	Y	Y	Y	Y
9	Y	Y	N	N	N	N	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y
10	N	N	Y	N	Y	Y	N	N	N	N	N	Y	N	N	Y	Y	Y	Y
11	N	Y	N	N	Y	Y	N	Y	N	N	N	Y	Y	N	N	N	N	N
12	N	N	N	N	N	N	N	N	N	N	N	Y	N	N	N	N	Y	Y
13	N	Y	N	N	Y	N	Y	Y	N	N	N	Y	Y	Y	Y	Y	Y	Y
14	N	Y	N	N	Y	N	Y	Y	N	N	N	Y	N	Y	N	Y	Y	Y
15	Y	Y	N	N	Y	N	Y	Y	N	Y	N	N	Y	Y	Y	Y	Y	Y
16	Y	N	N	Y	Y	N	Y	Y	N	Y	N	Y	N	N	N	Y	Y	Y
17	Y	N	N	Y	N	N	Y	Y	N	Y	N	Y	N	N	N	Y	Y	Y
18	Y	N	N	Y	Y	N	Y	Y	N	Y	N	Y	N	N	N	Y	Y	Y

3.5 Step V: Reachability Matrix and Transitivity check.

Y in the knowledge base cell is replaced by 1and N is replaced by 0 in reachability matrix. Check for transitivity and reachability matrix is constructed as shown in Table 3.

Table 3: Reachability matrix

Inhibitor No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
I	1	1	0	0	0	0	1	1	0	0	1	0	1	1	0	1	1	1
2	0	1	0	0	1	0	0	1	0	0	1	1	0	0	0	1	1	1
3	0	0	1	0	0	0	0	1	0	0	1	1	0	0	0	0	1	1
4	1	1	0	1	0	0	1	0	0	0	1	1	1	1	0	1	1	1
5	0	0	0	0	1	1	0	1	0	0	1	1	0	0	0	0	1	1
6	0	1	0	0	1	1	0	1	0	0	1	1	0	0	0	1	1	1
7	0	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1
8	1	0	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1
9	1	1	0	0	0	0	1	1	1	0	1	0	1	1	1	1	1	1
10	0	0	1	0	1	1	0	0	0	1	1	1	0	0	1	1	1	1
11	0	1	0	0	1	1	0	1	0	0	1	1	1	1	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1
13	0	1	0	0	1	0	1	1	0	0	1	1	1	1	1	1	1	1
14	0	1	0	0	1	0	1	1	0	0	1	1	0	1	0	1	1	1
15	1	1	0	0	1	0	1	1	0	1	1	0	1	1	1	1	1	1
16	1	0	0	1	1	0	1	1	0	1	0	1	0	0	0	1	1	1
17	1	0	0	1	0	0	1	1	0	1	0	1	0	0	0	1	1	1
18	1	0	0	1	1	0	1	1	0	1	0	1	0	0	0	1	1	1

3.6 Step VI: Level Partition in Reachability Matrix.

ISM based level partition is carried out. Reachability set and antecedent sets for all the elements are determined. Intersection of the two sets is found out. The elements for which the reachability set and intersection set remain same, occupy the top level in ISM hierarchy. Top level elements will not influence the remaining elements hence it can be removed from further calculation. The same process is repeated until the levels of each element are found out. Level partition details are shown in Table 4 to 12.

Table 4 - 1st Level partition

Inhibit No.	Reachability Set	Antecedent Set	Intersection	Level
1	1,2,7,8,11,13,14,16,17,18	1,4,8,9,15,16,17,18	1,8,16,17,18	
2	2,5,8,11,12,16,17,18	2,4,6,9,11,13,14,15	2,11	
3	3,11,12,17,18	3,5,7,10,11	3,11	
4	1,2,4,7,8,11,12,13,14,16,17,18	4,7,8,16,17,18	4,7,8,16,17,18	
5	5,6,8,11,12,17,18	2,5,6,7,8,10,11,13,14,15,16,18	5,6,8,11,18	
6	2,5,6,8,11,12,16,17,18	5,6,7,10,11	5,6,11	
7	3,4,5,6,7,8,11,12,13,14,15,16,17,18	1,4,7,8,9,13,14,15,16,17,18	4,7,8,13,14,15,16,17,18	
8	1,4,5,7,8,12,13,14,15,16,17,18	1,2,4,5,6,7,8,9,11,13,14,15,16,1	1,4,5,7,8,13,14,15,16,17,	

		7,18	18	
9	1,2,7,8,9,11,13,14,15,16,17,18	9	9	
10	3,5,6,10,11,12,15,16,17,18	10,15,16,17,18	10,15,16,17,18	
11	2,5,6,8,11,12,13,14	1,2,3,4,5,6,7,9,10,11,13,14,15	2,5,6,11,13,14	
12	12,17,18	2,3,4,5,6,7,8,10,11,12,13,14,16, 17,18	12,17,18	1
13	2,5,7,8,11,12,13,14,15,16,17,18	1,4,7,8,9,11,13,15	7,8,11,13,15	
14	2,5,7,8,11,12,14,16,17,18	1,4,7,8,9,11,13,14,15	7,8,11,14	
15	1,2,5,7,8,10,11,13,14,15,16,17, 18	7,8,9,10,13,15	7,8,10,13,15	
16	1,4,5,7,8,10,12,16,17,18	1,2,4,6,7,8,9,10,13,14,15,16,17, 18	1,4,7,8,10,16,17,18	
17	1,4,7,8,10,12,16,17,18	1,2,3,4,5,6,7,8,9,10,12,13,14,15, 16,17,18	1,4,7,8,10,12,16,17,18	
18	1,4,5,7,8,10,12,16,17,18	1,2,3,4,5,6,7,8,9,10,12,13,14,15, 16,17,18	1,4,5,7,8,10,12,16,17,18	

Table 5-2nd Level

Elements	Reachability Set	Antecedent Set	Intersection	Level
1	1,2,7,8,11,13,14,16	1,4,8,9,15,16	1,8,16	
2	2,5,8,11,16	1,2,4,6,9,11,13,14,15	2,11	
3	3,11	3,5,7,10,11	3,11	2
4	1,2,4,7,8,11,13,14,16	4,7,8,16	4,7,8,16	
5	5,6,8,11	2,5,6,7,8,10,11,13,14,15,16	5,6,8,11	
6	2,5,6,8,11,16	5,6,7,10,11	5,6,11	
7	3,4,5,6,7,8,11,13,14,15,16	1,4,7,8,9,13,14,15,16	4,7,8,13,14,15,16	
8	1,4,5,7,8, 13,14,15,16	1,2,4,5,6,7,8,9,11,13,14,15,16	1,4,5,7,8,13,14,15,16	
9	1,2,7,8,9,11,13,14,15,16	9	9	
10	3,5,6,10,11,15,16	10,15,16	10,15,16	
11	2,5,6,8,11,13,14	1,2,3,4,5,6,7,9,10,11,13,14,15	2,5,6,11,13,14	
13	2,5,7,8,11,13,14,15,16	1,4,7,8,9,11,13,15	7,8,11,13,15	
14	2,5,7,8,11,14,16	1,4,7,8,9,11,13,14,15	7,8,11,14	
15	1,2,5,7,8,10,11,13,14,15,16	7,8,9,10,13,15	7,8,10,13,15	
16	1,4,5,7,8,10,16	1,2,4,6,7,8,9,10,13,14,15,16	1,4,7,8,10,16	

Table 6-3rd Level

Elements	Reachability Set	Antecedent Set	Intersection	Level
1	1,2,7,8, 13,14,16	1,4,8,9,15,16	1,8,16	
2	2,5,8, 16	1,2,4,6,9,13,14,15	2	
4	1,2,4,7,8,13,14,16	4,7,8,16	4,8,16	
5	5,6,8	2,5,6,7,8,10,13,14,15,16	5,6,8	3
6	2,5,6,8, 16	5,6,7,10	5,6	
7	4,5,6,7,8, 13,14,15,16	1, 4,7,8,9,13,14,15,16	4,7,8,13,14,15,16	
8	1,4,5,7,8, 13,14,15,16	1,2,4,5,6,7,8,9,13,14,15,16	1,4,5,7,8,13,14,15,16	

9	1,2,7,8,9, 13,14,15,16	9	9	
10	5,6,10, 15,16	10,15,16	10,15,16	
13	2,5,7,8, 13,14,15,16	1,4,7,8,9,13,15	7,8,13,15	
14	2,5,7,8, 14,16	1,4,7,8,9,13,14,15	7,8,14	
15	1,2,5,7,8,10, 13,14,15,16	7,8,9,10,13,15	7,8,10,13,15	
16	1,4,5,7,8,10,16	1,2,4,6,7,8,9,10,13,14,15,16		

Table 7-4th Level

Elements	Reachability Set	Antecedent Set	Intersection	Level
1	1,2,7, 13,14,16	1,4,9,15,16	1,16	
2	2,16	1,2,4,9,13,14,15	2	4
4	1,2,4,7,13,14,16	4,7,16	4,7,16	
7	4,7,13,14,15,16	1,4,7,9,13,14,15,16	4,7,13,14,15,16	
9	1,2,7, 9, 13,14,15,16	9	9	
10	10, 15,16	10,15,16	10,15,16	
13	2,7,13,14,15,16	1,4,7,9,13,15	7,13,15	
14	2,7,14,16	1,4,7,9,13,14,15	7,14	
15	1,2,7,10, 13,14,15,16	7,9,10,13,15	7,10,13,15	
16	1,4,7, 10,16	1,2,4,7,9,10,13,14,15,16	1,4,7,10,16	

Table 8-5th Level

Elements	Reachability Set	Antecedent Set	Intersection	Level
1	1, 7, 13,14	1,4,9,15	1	
4	1, 4,7,13,14	4,7	4,7	
7	4, 7, 13,14,15	1,4,7,9,13,14,15	4,7,13,14,15	
9	1, 7, 9, 13,14,15	9	9	
10	10, 15	10,15	10,15	
13	7,13,14,15	1,4,7,9,13,15	7,13,15	
14	7,14	1,4,7,13,14,15	7,14	5
15	1,7,10, 13,14,15	7,9,10,13,15	7,10,13,15	

Table 9 -6th Level

Elements	Reachability Set	Antecedent Set	Intersection	Level
1	1,13	1,4,9,15	1	
4	1, 4,13	4	4	
9	1,9, 13,15	9	9	
10	10, 15	10,15	10,15	6
13	13,15	1,4,9,13,15	13,15	
15	1,10, 13,15	9,10,13,15	10,13,15	

Table 10-7th Level

Elements	Reachability Set	Antecedent Set	Intersection	Level
1	1,13	1,4,9	1	
4	1, 4,13	4	4	
9	1,9, 13	9	9	
13	13	1,4,9,13	13	7

Table 11-8th Level

Elements	Reachability Set	Antecedent Set	Intersection	Level
1	1	1,4,9	1	8
4	1, 4	4	4	
9	1,9	9	9	

Table 12.-9th Level

Elements	Reachability Set	Antecedent Set	Intersection	Level
4	4	4	4	9
9	9	9	9	9

3.7 Step VII: Diagraph development.

The elements are arranged graphically in levels and links are drawn as per the relationship. Diagraph is used to represent the elements and their interdependence in terms of nodes and edges. Elements are arranged graphically in levels and the directed links are drawn as per the relationship shown in reachability matrix. Diagraph is shown in Figure 1.

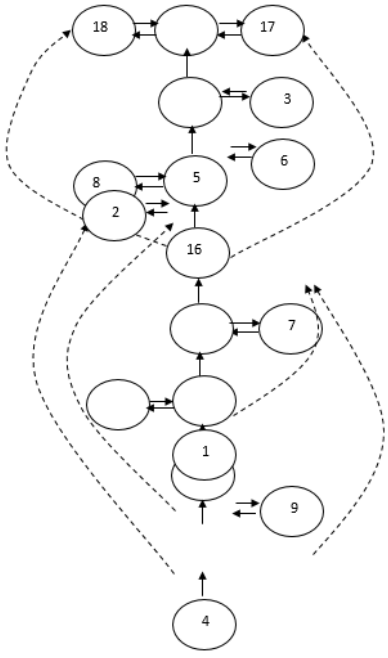


Figure 1: Diagraph with significant transitive links.

3.8 Step VIII: Interaction matrix.

The diagraph is translated into a binary interaction matrix form depicting all the interactions by 1 in cells. Remaining cell entry is 0. Cell with 1 is interpreted by picking the relevant interpretation from the knowledge base in the form of interpretations matrix. Interpretation matrix is shown in

Table 13. Explanation of interpretation matrix is as given below.

Table 13.-Interpretation Matrix.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0
2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
3	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
4	1	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
5	0	0	0	0	1	1	0	1	0	0	1	0	0	0	0	0	0	0
6	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0
8	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
9	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0
11	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1
13	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0
14	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0
15	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0
16	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
17	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0
18	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1

1-13—Lack of security will affect network level security in infrastructure. Infrastructure security at network level can be done by considering all sources in cryptography .Authorization and authentication can be assured for keeping security.

1-16-- Lack of sufficient security will affect data security. There is chance for intruders to attack database by introducing fraudulent data if there is lack of sufficient data security there by decreasing confidentiality of the customer.

2-16—Lack of data security can cause lack of reliability. If the data obtained from cloud is not correct then the reliability may decreases.

3-11---Network management barriers and lack of portability are related together. Lack of portability means the same software is not able to use in different countries. Portability of the cloud gets diminished due to network management barriers.

4-1—Lack of privacy leads to lack of sufficient security.

4-2—Privacy and reliability are related together.Lack of both will influence cloud services. If the customer is not reliable with cloud services then number of customers will get decreased which will ultimately affect the system.

4-9—Lack of privacy and inefficient backup management are affecting the cloud services. If having a good back up management then even the lost data can be reloaded.

5-6—Comprehensive management tool would help automatic service provisioning, balance workloads and aids with capacity planning and configuration management.

5-8—Lack of data confidentiality is fatal reason of the failure of cloud services. Customers should have very high confidentiality in the services of the cloud. Confidentiality can be increased by providing data security and data integrity.

5-11—Common standard for the cloud is necessary so as to improve portability. Changes in the security requirement may change the topology of network.

6-5—Automatic service provisioning and configuration management are considered. A better standard is necessary for the cloud computing services.

7-14—Weak access control is another problem of cloud computing. Infrastructure security at host level is one of the reasons of weak access control. By providing all securities in all levels of infrastructure this can be remedied.

8-5—Lack of confidentiality of customers may affect cloud services.

9-4—Ineffective back up management may cause serious problems if some data get lost unexpectedly.

9-7-- Access control can manage data, racking of servers, networking and cabling.

10-15—Cost/ time barrier is very essential for close watching of cloud services. Cost of running an industry without using cloud is expensive. Cloud migration is more feasible, realistic and less expensive.

11-12—Network management barriers indirectly cause legal issues.

12-17—some customers depends cloud for storing data. If data integrity is not sustained it may lead to legal issues.

12-18—Lack of data availability make customers feel bad about the services of cloud.

13-7—Infrastructure security at network level should be considered seriously and it may cause weak access control.

13-15—Infrastructure security at network level and application level are very important. Correct measurement should be taken for keeping security in both levels.

14-16—Infrastructure security at host level influence lack of data security.

15-10—Infrastructure security at application level should be considered very serious. Cost/time barrier may depend on it.

16-2—Lack of data security is directly proportional to lack of reliability. Reliability can be increased by providing good security for data in cloud.

17-12—Lack of data integrity may lead to legal issues.

18-12—Lack of data availability can also end in legal issues.

3.9 Step IX: Total interpretive structural model.

Total Interpretive Structural Model is obtained from interpretive matrix and diagraph .The nodes are replaced with boxes having elements. Interpretation is depicted on the side of the links. The Total Interpretive Structural Model is shown in Figure 2.

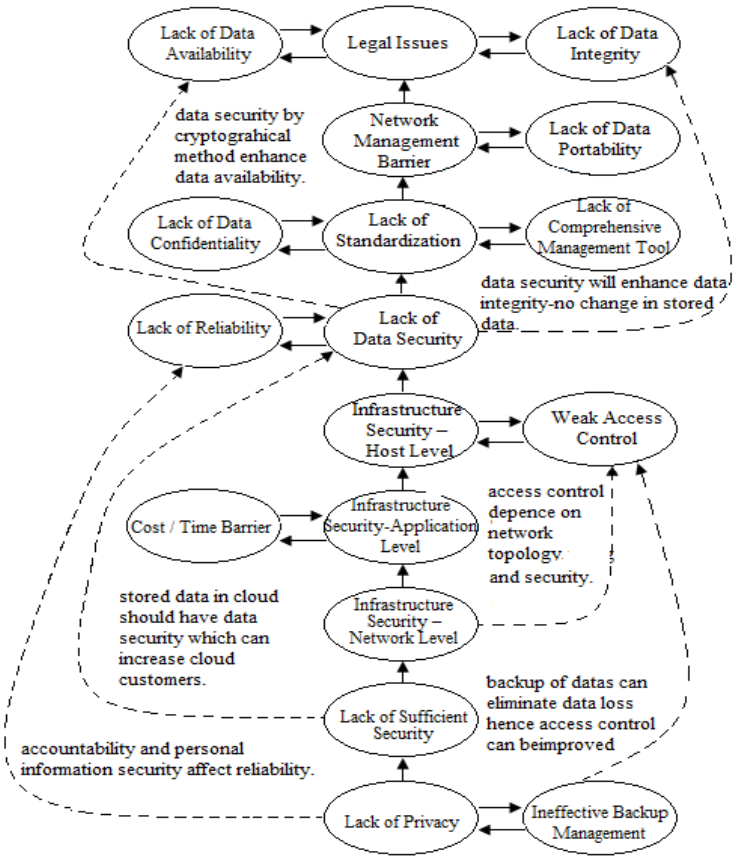


Fig: 2 -Total interpretive structural model

4 Results and discussions

Inhibitors are put in different levels so as to get interrelationship of inhibitors. Level partition is done by considering reachability set, antecedent set and intersection set as described in the section. There are nine Partition levels. The elements coming under each level is discussed below.

The interrelationships of inhibitors are found out from the diagraph. First level elements are legal issues, lack of data integrity, and lack of data availability. The second level elements are network management barriers and lack of portability. Third level elements are lack of standards, lack of confidentiality, and lack of comprehensive management tools. Lack of data security and lack of reliability are under fourth level. Infrastructure security at host level and weak access control are in the fifth level. Infrastructure security at application level and cost/time barrier are in sixth level. Infrastructure security at network

level is coming under seventh level. Lack of sufficient security is coming under eight levels. Lack of privacy and inefficient backup management are in ninth level. Interrelationship among them is shown in the Figure 1. Lack of privacy is very essential and cloud service providers should have to take precautions for it. Inefficient backup management is directly related to weak access control. Lack of privacy is related to lack of reliability, lack of Confidentiality has an indirect relation to data availability. Lack of sufficient security is related to data security. Alidades in cloud should be protected by providing modern concepts of crypto graphical protection. Lack of data security will be directly related to data integrity. Lack of data security affects data availability. Confidentiality level of the customers can be increased by providing data availability and data integrity. Legal issue is related to lack of data integrity and lack of data availability.

5 Conclusion

Inhibitors of cloud computing are studied and their interrelationship is figured out. Interrelationship studied by using ISM has no interpretation for the relation whereas Total Interpretive Structural Modeling has interpretation of the relation. In TISM logic behind the interrelation is clarified through the expert's opinion. Contextual relationship in SSIM remains silent in ISM whereas in TISM the real working is considered. TISM of inhibitors of cloud computing is drawn and the relationship is explained in interaction matrix. Major inhibitors should be considered before going for cloud installation.

REFERENCES

[1]. Tim Mather, SubraKumaraswamy, ShahedLatif, "Cloud Security and Privacy- An Enterprise Perspective on Risks and Compliance" O Reilly

[2]. Wang.C and Wulf W. A., (1997,) "Towards a framework for security measurement", 20th

[3]. National Information Systems Security Conference, Baltimore, MD, pp. 522-533.

[4]. Savola.R and Abie.H, (2010)."Development of measurable security for a distributed

[5]. Messaging system," International Journal on Advances in Security, Vol. 2.

[6]. Jaquith. A, (2007)"Security metrics: replacing fear, uncertainty and doubt,"Addison-Wesley.

[7]. Gadia.S, (2009) "Cloud computing: an auditor's perspective," ISACA Journal, Vol. 6,.

[8]. Gellman.R,(2009) "Privacy in the clouds: risks to privacy and confidentiality from cloud computing," World Privacy Forum (WPF) Report.

[9]. Cloud Security Alliance, (2010) "Top threats to cloud computing", Version 1.0. Downloaded from: www.cloudsecurityalliance.org

[10]. Cloud Security Alliance. www.cloudsecurityalliance.org [July 4, 2010].

[11]. Mandal.A, Deshmukh.S,(1994) Vendor selection using interpretive structural modeling (ism). International Journal of Operations and Production Management, , 14(6): 52–59.

[12]. Sage.A, (1977). Interpretive Structural Modeling: Methodology for Large-scale Systems, 91–164. McGraw-Hill, New York,

[13]. Warfield.J. (2005) Developing interconnection matrices in structural modeling. IEEE Transactionsons on Systems, Man and Cybernetics, 4(1): 81–67.

[14]. Wang.C(2009)"Forrester: A close look at cloud computing security issues," http://www.forrester.com/securityforum

[15]. IDC, "It cloud services user survey, pt.2: Top benefits & challenges," http://blogs.idc.com/ie/?p=210, 2008.

[16]. Zetta, (2008.) "Zetta: Enterprise cloud storage on demand," http://www.zetta.net/,

[17]. Chen.P, Lee.E, Gibson.G, Katz.R, and Patterson.D (1994.) "RAID: High performance,reliable secondary storage," ACM Computing Surveys (CSUR), vol. 26, no. 2,pp. 145–185

[18]. Yahoo!,"Hadoop distribted file system architecture," http://hadoop.apache.org/common/docs/current/hdfs design.html, 2008.

[19]. Dwork.Cet al., "Differential privacy," LECTURE NOTES IN COMPUTER SCIENCE, vol. 4052, p. 1, 2006.

[20]. Dwork.C, "Differential privacy: A survey of results," Lecture Notes in Computer Science, vol. 4978, p. 1, 2008.

[21]. Dean. J and Ghemawat.S,(2004), "MapReduce: simplified data processing on large clusters," in Proceedings of the 6th conference on Symposium onOpearting Systems Design & Implementation-Volume 6 table of contents, pp. 10–10.

[22]. Bardin, (2009.) J"Security Guidance for Critical Areas of Focus in Cloud Computing," www. cloudsecurityalliance.org/ guidance/ csaguide.pdf,

[23]. Hwang, K G. Fox, and Dongarra.J,(2010.) Distributed Systems and Cloud Computing: Clusters, Grids/P2P, and Internet Clouds, Morgan Kaufmann, to appear,

[24]. Nick J,(2010.) " Journey to the Private Cloud: Security and Compliance," tech. presentation, EMC, Tsinghua Univ.,

[25]. Rittinghouse J and Ransome.J, (2010),Cloud Computing: Implementation, Management and Security, CRC Publisher,

[26]. "Gartner Says Cloud Computing Will be as Influential As E-business". Gartner.com. Retrieved 2010-08-22.

[27]. Ravi.V. and Shankar. R. (2005), Analysis of interactions among the barriers of reverse logistics, Technological Forecasting and Social Change, 72(8): 1011-1029.

[28]. Thakkar. J.,Kanda.A. and Deshmukh, S.G.(2008), Interpretive Structural Modeling (ISM) of IT-enablers for Indian manufacturing SMEs', Information management & Computer Security, Vol. 16 No.2, pp. 113-136

[29]. Quan Liu, Lu Gao, Ping Lou, "Resource Management Based on Multi-Agent Technology for Cloud Manufacturing, IEEE 2011

[30]. FarzadSabahi, "Cloud Computing Security Threats and Responses" IEEE, 2011.

[31]. Craig A Lee, " A Perspective on Scientific Cloud Computing", ACM 2010

[32]. P. Sasikala,(2011), " Cloud Computing: present status and the future implications", Inderscience Enterprises Ltd.

TNC Transactions on Networks and Communications
SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
VOLUME 2, ISSUE 5
ISSN: 2054 - 7420

A Review Study on Analytical Estimation of Optimal Number of Clusters in Wireless Sensor Networks

Vinay Kumar [1], Sanjay B. Dhok[2] , Rajeev Tripathi[3] and Sudarshan Tiwari[4]
[1,2]*Department of Electronics Engineering, Visvesvaraya National Institute of Technology, Nagpur, Maharashtra, India,*
[3]*Department of Electronics & Communication Engineering, Motilal Nehru National Institute of Technology, Allahabad, Uttar Pradesh, India,*
[4] *National Institute of Technology, Raipur, Chhattisgarh, India*
vk@ece.vnit.ac.in; sanjaydhok@gmail.com; rt@mnnit.ac.in; stiwari@mnnit.ac.in

ABSTRACT

To improve energy efficiency, total network scalability and data aggregation in Wireless Sensor Networks (WSNs), sensor nodes are often grouped into disjoint and mostly non-overlapping subsets called clusters. To provide an energy-efficient system by clustering, there are three main challenges. First is to find the optimum number of nodes in a specific cluster, second is to find the optimum number of clusters in the network and the third is to find the optimum position of Cluster Head (CH) in a specific cluster. Selecting an optimum number of clusters in WSNs provide greater improvement in terms of system scalability, energy efficiency, collision reduction, network lifetime, latency, and efficient routing backbone in the network. Selection of optimal number of clusters in WSNs is affected by level at which WSNs is modeled viz. Radio Energy Model Level, Network Model Level and Clustering Level. The objective of this paper is to present a state-of-the-art survey of distinct analytical methods used to calculate the optimum number of clusters, and its time-line comparative analysis based on network type, mathematical formula for an optimal number of clusters, base station positioning, energy model, strengths, weaknesses and applications of WSNs. We have also discussed the impact of different parameters on selecting the optimal number of clusters in WSNs.

Keywords: Wireless Sensor Networks, Clustering, Optimal Clustering, Energy Efficient WSNs, Optimal Number of Clusters, Algorithms for Optimal Number of Clusters

1 Introduction

WSNs are large-scale networks of small embedded devices, each with sensing, computation and communication capabilities and have been widely discussed in [1, 2, 8, 20]. In WSNs, sensor nodes have limited processing power, communication bandwidth, and storage space, which demand very efficient resource utilization. The sensor nodes are often grouped into individual disjoint sets called cluster [3, 4]. Clustering is used in WSNs [5, 6], as it provides network scalability, resource sharing and efficient use of constrained resources that give network topology stability and energy-saving attributes. Clustering schemes, offer reduced communication overheads, and effective resource allocations thus decreasing the overall energy consumption and reducing the interferences among sensor nodes. In sensing field if

DOI: 10.14738/tnc.25.488
Publication Date: 03rd October, 2014
URL: http://dx.doi.org/10.14738/tnc.25.488

we have more number of clusters while maintaining the same load per Cluster Heads (CHs), the communication distance from a sensor node to its own CH is reduced. Therefore, the overall energy consumption is also reduced. On the other hand, increasing the number of clusters means that the communication path between a sensor and the Base Station (BS) will include more cluster heads to the cluster head hops, which mean higher overall energy consumption. Accordingly, finding the optimal number of clusters is a very crucial point in the system [57]. Selecting an optimal number of clusters in WSNs provide greater improvement in terms of energy efficiency, system scalability, network lifetime, and latency. Cluster optimization does not play a significant role for moderate size sensor networks if free space fading energy is low, but for large networks, cluster size optimization is still important even if free space fading is low [51]. The optimal number of clusters is very sensitive to energy model and sensing model of the nodes used in the system. The objective of this paper is to present a state-of-the-art survey of distinct analytical methods used to calculate the optimal number of clusters and its comparison based on, network lifetime, expression for an optimal number of clusters, base station position, energy model used, advantages, disadvantages and applications in WSNs. We have also discussed the impact of different parameters on selecting the optimal number of clusters in WSNs**.**

Though there are number of survey papers on the topics of WSNs and clustering in WSNs, but none of the research papers surveyed algorithms for an optimal number of clusters analytically or theoretically in WSNs. To the best of our knowledge, this is the first survey paper which carried out a review study on different algorithms for finding optimal number of clusters analytically in WSNs. The rest of the paper is organized as follows: in section 2, we provide a basic idea for clustering, cluster characteristics and the need for clustering in WSNs. Section 3 presents optimal clustering & factors effecting the optimal number of clusters in WSNs. Section 4 presents a survey on state-of-art of different algorithms for finding the optimal number of clusters analytically, reported in the literature. The open issues and challenges in WSNs are discussed in 5 and finally paper is concluded in section 6.

2 Clustering, Cluster Characteristics and the need for Clustering in WSNs

2.1 Clustering in WSNs

To support high scalability and better data aggregation, sensor nodes are often grouped into disjoint and mostly non-overlapping subsets called clusters. In the clustering, each cluster has a leader, which is called the CH and it performs the tasks like fusion and aggregation of data. Figure 1 represents clustering in sensor networks along with inter cluster and intra cluster communications, CHs and sensor nodes**.**

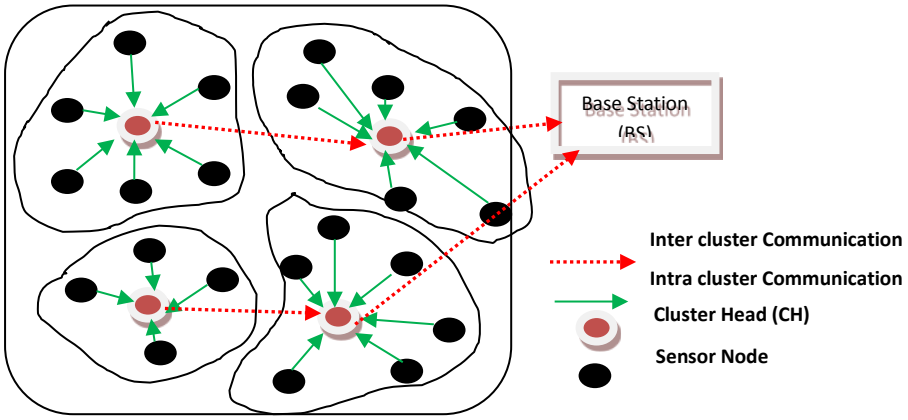


Figure 1: Clustering in WSNs with inter and intra cluster communication

The sensor nodes in a particular cluster periodically transmit their data to the CH nodes. CH nodes aggregate the data and transmit to the Base Station (BS) either using single-hopping or multi-hopping communication. The communication between CH and nodes is called intra-cluster communication and communication between CHs and base station is called inter-cluster communication [9]. The definitions of some terms are as follows: **Cluster head (CH):** CH aggregates the data sensed by the cluster members (sensor nodes) in a particular cluster and aggregated data will be transmitted to BS. **Base** station (BS): It has high processing capabilities and high level of energy. BS is the co-ordinator of the network where all the aggregated data from CHs are processed. **Sensor *node*:** Most of the nodes in the network, which are neither CHs nor BS, are considered simple sensor nodes.

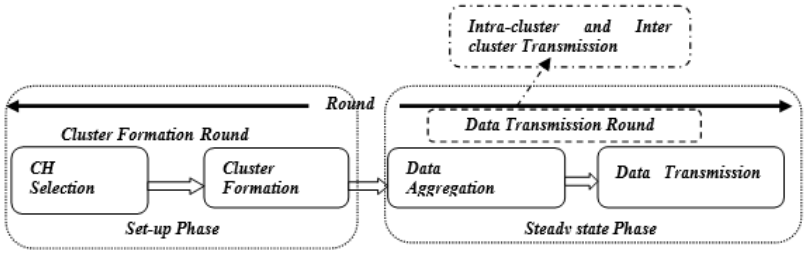


Figure 2: Phases in single round of clustering techniques

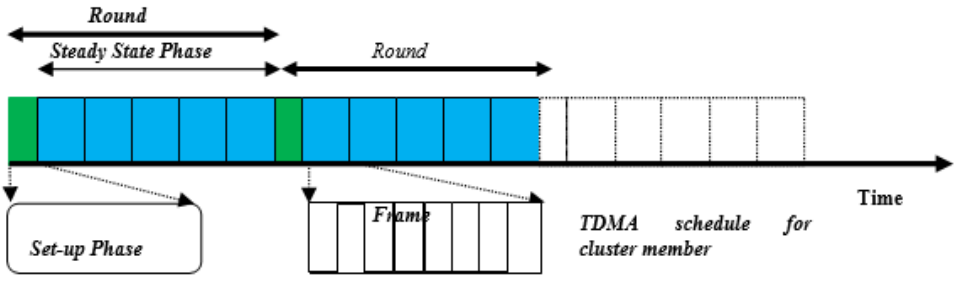


Figure 3: Timing diagram of phases in single round of clustering techniques

Cluster-based algorithms work in four stages: CH selection, cluster formation, data aggregation and data transmission. Most of the clustering protocols, divide the schedule of the network into different rounds of fixed duration. Figure 2, shows that each round consists of a setup phase and steady-state phase. During the set-up phase some sensor nodes elect themselves as CHs. The steady state phase, which is

sub divided into data aggregation and data communication. Steady state phase divided into different frames. During the steady-state phase, within each frame the cluster heads receive sensor data from cluster members (according to some multiple access technique and MAC protocol), and transfer the aggregated data to the BS [15]. Figure 3 represent timing diagram of phases in single round of clustering techniques.

2.2 Cluster characteristics

Cluster characteristics are very important in the clustering process of WSNs. Figure 4 shows the taxonomy of cluster characteristics in WSNs. Some cluster characteristics are defined as follows: **Cluster Changeability***:* Clustering techniques can be classified into two types: fixed and variable ones. In the fixed techniques, the set of CHs are predetermined and the number of CHs is fixed. In a variable technique number of CHs is variable, in which CHs are selected randomly, from the deployed sensor nodes [15]. **Cluster Sizes:** It can be classified into two types: uniform (same size clusters) and non-uniform (different size clusters), in the network. **Intra-Cluster connectivity**: This characteristic classified based on basis of communication inside a particular cluster; it includes two classes: single-hop and multiple-hop intra cluster connectivity and **Inter-Cluster Connectivity**: this characteristic classified on the basis of communication between the base station and cluster heads; its include two classes namely single hop and multi-hop inter cluster communication.

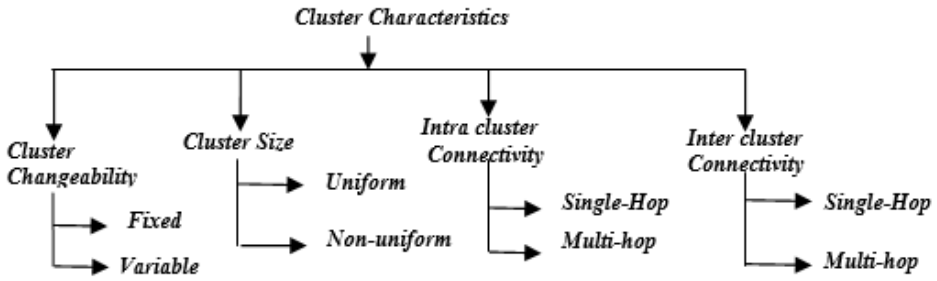


Figure 4: Taxonomy of Cluster characteristics

2.3 Why clustering in WSNs?

1. The advantages of clustering are many and are listed as follows:
2. It enables bandwidth reuse, thus can improve the system capacity within a cluster.
3. All the normal nodes send their data to the CHs so that energy saving is achieved by absence of flooding, multiple routes, routing loops [9, 14]
4. Clustering enables efficient resource allocation and thus helps in better designing of power control. Clustering facilitates data aggression/data fusion**.**
5. Any changes of node's behavior within a cluster affect only that cluster, but not the entire network, which make it robust to these changes [25].
6. Since the backbone network consists only the CHs, which are fewer in number than all the sensor nodes in the entire network. It requires less storage of routing information.
7. Clustering schemes make it easier in responding to changes caused by network dynamics, node mobility, unpredicted failures and local changes. Since these changes need to be managed and detected within individual clusters only [11, 12].

8. If the sensor nodes are mobile in nature, then nodes residing in concern clusters only need to update the information. Thus, local changes need not be updated by whole network, and this reduces the information processed and stored by each mobile sensor node [13].
9. In clustering, only CHs are responsible for transmitting data to base station this will reduce data collisions between the nodes [77, 78].
10. Generally, sensor network uses multi-hoping to transfer data to the BS. In this, the traffic transmitted by each node includes both self and relayed data. The sensor nodes closer to the BS have to transmit more data than those far away from the BS. As a result, the nodes closest to the BS heavily deplete their energy, creating a hole near the BS. So with the help of clustering hole problem can be reduced [17, 18, 19,79].

3 Optimal Clustering in WSNs

The main idea behind optimal clustering (selecting the optimal number of clusters or Cluster Heads) is to determine a clustering of the network such that the entire energy required for collecting data from the whole network is minimized as compared to other possible clustering patterns [67].

3.1 Why Optimal Clustering in WSNs?

If the clusters are not constructed in an optimal way, the total energy consumed by the sensor network per round is increased exponentially when the number of clusters created is greater especially when the number of the constructed clusters is less than the optimal number of clusters [30].

i. In sensing field, choosing more clusters while maintaining the same load per Cluster Heads (CHs), the communication distance from a sensor node to its own CH is reduced. Therefore, the overall energy consumption is also reduced. On the other hand, increasing the number of clusters means that the communication path between a sensor and the BS will include more CH to CH hops, which mean higher overall energy consumption. Therefore, finding the optimal number of clusters is a crucial point for the WSNs [57].
ii. In WSNs from the ***Physical (PHY) layer*** point of view, using a large number of clusters can reduce energy consumption because the communication distance between CHs can be reduced. From the ***Medium Access Control (MAC)*** layer point of view, using a lesser number of clusters can reduce energy consumption because it decreases the average possibility of being a cluster head for each sensor node and from ***Network Layer*** point of view, lesser clusters yield fewer hop counts to the data sink, and result in less energy consumption. So for a cross-layer trade-off design issue among the required communication power in the physical layer, the possibility of being CHs in the MAC layer and the number of hops in the relay path in the network layer. So we have to optimize the number of clusters in WSNs [53].
iii. Consider the possibility of processing data inside the cluster. After processing of the data, it will be transferred to CHs. Energy consumption decreases with increasing cluster sizes because data traffic decreases and data aggregation rate grows. However, for very large cluster, the performance is rather irrelevant, so optimal number of clusters should be selected [76].

With the help of above given statements, we can say that optimal clustering in WSNs plays a great role. It provides benefits like limited resources can be utilized more efficiently, overall energy efficiency is improved and sensor network lifetime is improved.

3.2 Parameters affecting an optimal number of clusters at different levels in WSNs

The parameters affecting an optimal number of clusters in WSNs are divided into three levels: Radio model level, Network level and clustering level. Table 1 represents all the factors affecting the optimal number of clusters in WSNs.

Table 1: Parameters Affecting Optimal number of clusters at different levels in WSN

Level	Parameters
Radio Model Level	Energy Models
	Sensing Model
	Shadowing and path loss Exponent
Network Level	Node Density
	Transmitter and Receiver Circuit
	Size of Sensing Fields
	Number of Base Stations
	Base Station Positioning
	Single and Multi-Hopping
Clustering Level	Distance Between Cluster heads and
	Position of Cluster Heads
	In-network Processing
	Data Correlation and Distortions

The parameters affecting optimum number of clusters are described below:

3.2.1 Radio Model Level Parameters

i. **Energy Model:** Optimal number of clusters depends highly on the type of energy model used. Therefore, it is important to use the right energy model. In WSNs there are four energy model which are generally used [10, 23, 24, 25]. Most of the surveyed algorithms in this paper using energy model represented and explained by figure 5 [10].

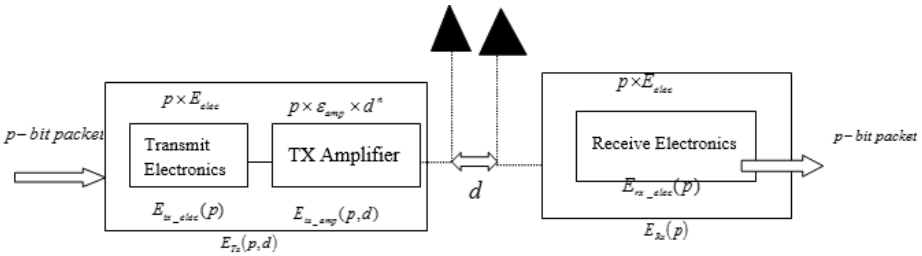


Figure 5: Radio Energy Dissipation Model

The energy consumption model can be simplified for a transmitter–receiver pair at distance d apart as Follows $E_c = E_{Tx}(p,d) + E_{Rx}(p)$. Where $E_{Tx}(p,d)$ and $E_{Rx}(p)$ are the energy consumption of the transmitter and the receiver, respectively. $E_{elec} = E_{tx\_elec}(p) = E_{rx\_elec}(p)$ =Energy dissipated to run the transmitter or the receiver circuitry to transmit or receive one bit of the data packet.

$\varepsilon_{amp} = E_{tx\_amp}(p,d)$ is energy dissipation of the transmission amplifier to convey one bit of data packet to the receiver node at distance of d=1m away

The energy consumption at the transmitter is divided into the transmit electronics and transmitter amplifier while the receiver energy consumption depends only on the receiver electronics. Then, the transmitter and receiver energy consumptions are:

$$E_{Tx}(p,d) = pE_{elec} + p\varepsilon_{amp}d^n \text{ and } E_{Rx}(p) = pE_{elec}$$

where $p =$ Length of transmitted/ received message in bits, d=distance between transmitter and receiver node, n=path loss exponent.

n=2 for free space model ($\varepsilon_{amp} = \varepsilon_{fs}$, when $d < d_0$) and n=4 for Multipath Model ($\varepsilon_{amp} = \varepsilon_{mp}$ when $d > d_0$)

$$l\varepsilon_{fs} \times {d_0}^2 = l\varepsilon_{mp} \times {d_0}^4 \qquad d_0 = \sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{mp}}}$$

Table 2: This data is well taken from [51], optimal number of clusters with different energy models, for 100 numbers of sensor nodes(N) and side of sensing area M= 100, when the distance between the CH and the sink node is between 45–145m.

Energy Model	K_{opt}	
	$\varepsilon_{fs} = 7nJ/bit/m^2$	$\varepsilon_{fs} = 10pJ/bit/m^2$
	N=100, M=100	N=100, M=100
Halgamuge et al.[51]	1-6	0-2
Zhu et al.[24]	1-12	0-4
Mille et al.[23]	2-13	0-5
Heinz et al.[31]	2-16	1-7

The difference in the optimal number of clusters (K_{opt}) between these energy models is getting closer as the distance between the sink and the CHs increases, This is because, as this distance increases the energy dissipation for communication becomes more and more dominant in the cost function.

ii. **Sensing Model:** It affects the optimal number of clusters because the distance between the cluster head and base station change with change of sensing model of the nodes. Basically there are two types of sensing models reported in literature. The first model is deterministic, and the second model is probabilistic **(**Boolean sensing model, shadow-fading sensing model and Elfes sensing model). According to this model [26, 27], the probability that a sensor detects an event to a distance χ

Where R_1 defines the starting of uncertainty in sensor detection and the parameters ψ *and* υ are adjusted according to the physical properties of the sensor. R_{max} is the maximum sensing range of the node

$$p(x) = \begin{cases} 1 & x \le R_1 \\ Exp(-\psi(x - R\ )^{\nu}) & R_1 \le x \le R_{Max} \\ 0 & x \ge R_{\max} \end{cases} \quad (1)$$

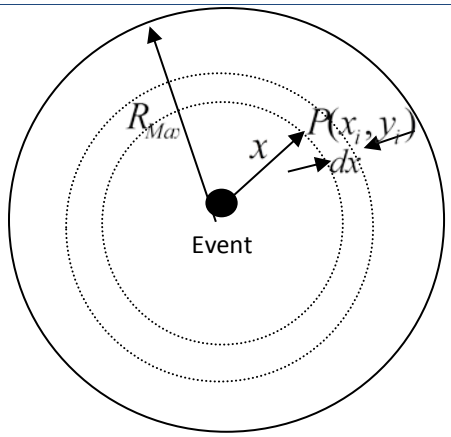


Figure 6: Probabilistic sensing model

With the help of above given probability density function one can find the distance between the base station and cluster heads. Figure 6 represents the probabilistic sensing model.

iii. **Shadowing and Path Loss Exponent Model:** The impact of the shadowing effect on optimal number of clusters is more significant for a larger value of the path loss exponent (*n)* than that for a smaller *n*. A larger *n* may amplify the shadowing effect on the optimal number of clusters [53, 62].

3.2.2 Network Level Parameters

i. **Node Density:** Experimental results show that for lower densities (275-375 nodes/km^2 approximately) the optimal cluster size is large and for higher densities (400-500 nodes/km^2 approximately) the optimal cluster size is one-hop when sensing field size is fixed. This is due to total intra cluster communication overhead for low density nodes [57].

ii. **Transmitter and Receiver Circuit Energy:** The energy consumption of the transmitter circuitry E_{tx_elec} has no impact on the optimal number of clusters. The energy consumption of the receiver electronics E_{rx_elec} can greatly change the optimal number of clusters. It is a very important factor that can decide on whether or not it is worth performing clustering in the sensor network [62].

iii. **Size of Sensing Field:** The optimal number of clusters can be independent of the sensing field size under the following conditions [7, 62]:

- ✓ E_{rx_elec} is small compared to E_{tx_elec}
- ✓ The wireless transmissions are governed by free space radio propagation model
- ✓ The BS is not located outside of the sensing field**.**

If the above requirements are satisfied, the optimal number of clusters will only depend on the number of sensor nodes across the network $K_{opt} = f(N)$. If the sensing area is square-shaped or circular, the optimal number of clusters can be expressed as $K_{opt} = \tau\sqrt{N}$ where τ is a constant with maximum value is one. Table V shows the impact of sensing field, base station position, path loss model on selecting the optimal number of clusters, which support the justification of the above statements.

iv. **Number of base stations:** Inter cluster communication increases when the number of base stations increases. For large network the energy consumption is nearly the same for single and two hop clusters. However, with three base stations and large networks, the optimal cluster size is 2 hops for reducing energy consumption in the network [57]**.**

v. **Base Station Positioning**: Optimal number of clusters will be larger when BS is located in the center of the sensing field. As BS moves away from the center of the sensing field area towards boundary to outside of sensing field, optimal number of clusters will be reduced.

vi. **Single and Multi-hopping:** The impact of selecting the optimal cluster size on total energy consumption is more prominent in single-hop communication than in a multi-hop WSNs. This is because the energy function is proportional to the square of the distance over which data transmission is done and this distance for the single-hop communication is often longer than that for the multi-hop case. When the degree of data correlation is high, then the wider range of cluster sizes will occur. This is also the reason why the optimal number of clusters in multi-hop communication gets larger than that of single hop approach as data correlation degree increases [67]. Single-hop clustering performs best for a large spectrum of different size of sensing field, node densities and the number of base stations. For very high density networks (more than 1000 nodes), multiple base stations (more than three) or very low density network (less than 400 nodes) 2-hop clustering performs better [57].

vii. **Free Space Fading Energy:** Selecting a number of clusters does not play a key role for reasonable size sensor networks if free space fading energy is low. But for large-scale networks, finding the number of clusters is still important, even if free space fading is low [57].

3.2.3 Clustering Level Parameters

i. **Distance between Base station and CHs*:*** Selection of optimal number of clusters also depends upon the distance between CHs and BS. The distance between the cluster head and base station depends upon the size of sensing field and type of sensing models.

ii. **Position of the cluster head:** The Optimal number of clusters will be large when BS is located in the center of the sensing field. As BS moves from the center of the sensing field area towards a boundary to outside of *sensing field, optimal number of clusters will be reduced. The position of the CH is also important because of two* reasons. First, it affects the load balance, and therefore energy consumption inside the cluster for data routing aggregation. Second, the overhead during routing of data inside the CHs with the head placement [57].

iii. **In Network Processing**: Consider the possibility of processing data inside the cluster. After processing of the data, data will be transferred to CHs. Energy consumption decreases with increasing cluster sizes because of the fact that data traffic decreases and data aggregation rate grows. However, for very large clusters the performance is rather irrelevant; therefore preference should be given to 3-4 hop clusters since they have simultaneously low energy consumption and low data aggregation rate [57].

iv. **Data correlation and distortion:** Finding the optimal cluster size depends on the value of the correlation. A large cluster size is optimal for low correlation and a small cluster size performs optimally for high correlation, there exist intermediate cluster sizes that perform near optimally over a wide range of spatial correlations. This near-optimal cluster size depends only on base station position and total number of nodes in the sensing field [36, 67].

4 Algorithms for analytically estimating the optimal number of clusters in WSNs:

There have been several different criteria to initially classify the algorithms for finding optimal number of clusters in WSNs. Two of the most common classifications are the algorithms for homogenous sensor networks the algorithms for heterogeneous sensor networks. Both classifications are based on the characteristics and functionality of sensor nodes in the cluster. Figure 7 shows the taxonomy of various algorithms for an optimal number of clusters. Table 3 represents symbols and their meaning used in different algorithms surveyed. Table 4 represents comparative analysis of algorithms based on network type, analytical value of optimal number of clusters and base station positioning. Table 5 presents comparative analysis of algorithms based on sensing field, radio model, base station position and analytical value of optimal number of clusters particularly for [62, 81, and 82]. Table 6 presents timeline comparative analysis of algorithms based on cluster variability, node distribution, energy model and type of sensing fields and finally table 7 presents comparative analysis of algorithms based on their strengths, weaknesses and applications.

4.1 Homogeneous sensor networks

A homogeneous sensor network consists of identical sensor nodes. It means that the sensor nodes have the same energy and hardware complexity. For static clustering in a homogeneous network, it is shown that the cluster head (CH) nodes will be over-loaded when long range transmission is required to transmit data to remote base station. In a heterogeneous sensor network, the nodes are enabled with extra battery energy and extra hardware. Sensor nodes in a particular cluster, use multi-hopping to send data to the respective cluster head. The sensor nodes that are closest to the cluster head have the highest energy burden due to relaying. When the sensor nodes use single hopping to transfer data to the cluster head, the sensor nodes that are farthest from the cluster heads always spend more energy than the sensor nodes that are closer to the cluster heads. Non-uniform energy consumption in the network takes place. Thus, there are two desirable characteristics of a sensor network: minimum hardware cost, and uniform energy consumption. While heterogeneous networks achieve the minimum hardware cost, the homogeneous networks achieve uniform energy [70, 83]. Most of the known algorithms for finding optimal number of clusters for WSNs can be further distinguished into two main parts: probabilistic (random) and non probabilistic, depending on the cluster formation method and the parameters used for CH selection. Optimal Probabilistic algorithms are further divided into two parts based on node distribution: uniform and non-uniform.

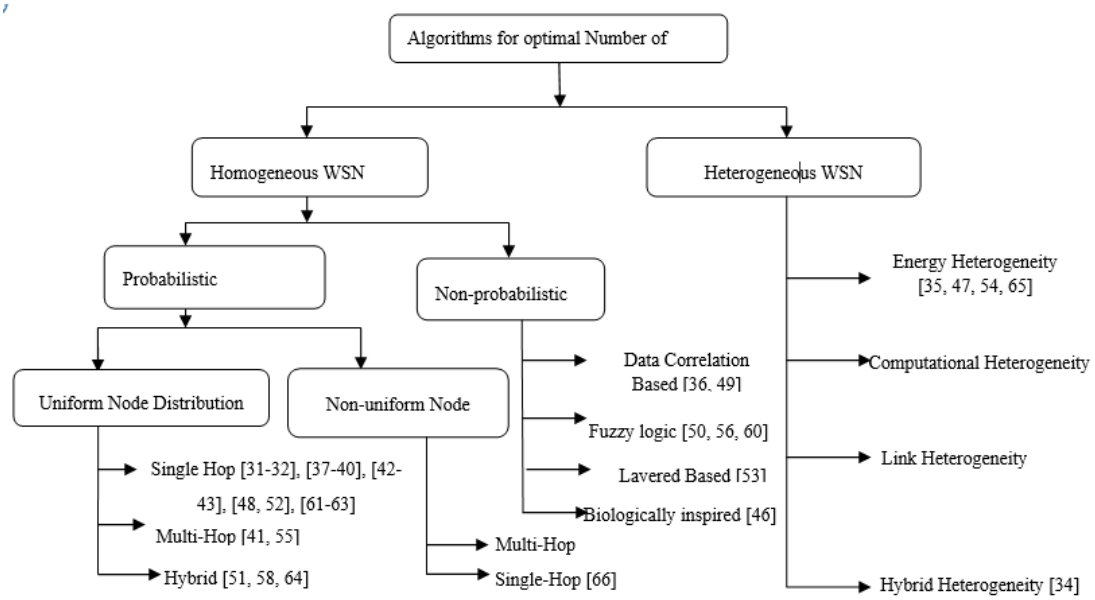


Figure 7: Taxonomy for the optimal number of clusters in WSNs

4.1.1 Probabilistic and Uniform Node Distribution

The main aim of probabilistic algorithm is to reduce the energy consumption and prolong the networks lifetime. Some of the algorithms follow a random approach for CH selection, whereas other follows a hybrid probabilistic model for CH selection.

Heinzelman et al. proposed [31], Low Energy Adaptive Clustering Hierarchy (LEACH) protocol, is the first and the oldest algorithm for the calculation of optimal number of clusters in WSNs. It uses the following techniques to achieve the design goals: randomized, self-configuring and adaptive cluster formation, local control for data transfers and low-energy media access control and application-specific data processing. LEACH protocol has many rounds, and each round has two phases, a setup phase and steady-state phase. In the setup phase; it provides cluster formation in an adaptive manner, and in the steady-state phase transfer of data take place. LEACH uses a TDMA/CDMA MAC to reduce inter-cluster and intra-cluster collisions. Cluster formation is based on many properties such as the number and type of sensors, communication range and geographical location [80]. Depedri et al. [32] proposed decentralized algorithms for cluster formation in which sensor node only knows about its own position and position of final receiver and not the position of all sensor nodes. It operates in following phases: Cluster head selection algorithm, Cluster formation and data transmission with multiple accesses. Each sensor node chooses its cluster head by evaluating the energy dissipated in the path between final receiver and itself. It provides better energy efficiency than LEACH. In this, each sensor can determine its own optimum number of clusters. This value depends on the total number of sensors in the network (N), path-loss exponent (n), the dimension of sensing field. It does not depend on the distance between the cluster head, and the base station like instead was in original LEACH [31].

Chen et al. [39] have determined two aspects of the problem: first for given the number of clusters, how does one choose the CHs to cover up the sensor network more efficiently? Second, how does one assess the number of clusters needed to utilize data correlation of sensors for a general sensor network? Comeau et. al. [38] have proposed multi-level clustered network based on single-hop communication.

Authors have analytically determined the optimal number of clusters at each level. Kim et. al. [37] has estimated the optimal number of clusters among random distributed sensors in a bounded sensing field. In this algorithm, optimal number of CHs depends on the distance between the base station and sensor node. Vlajic et al. [40] have proposed that in under some specific condition clustered WSNs provide an advantage over non clustered WSNs. This algorithm considerably more efficient in terms of energy conservation than clustering algorithms previously proposed in the literature.

Wang et al. [43] proposed clustering algorithm, which provides fixed optimum number of clusters in each round during the whole lifetime of the network. In LEACH protocol, optimum number of clusters in each round depends on the total number of nodes in the network, but theoretically, the total number of sensor nodes decreased and the optimal number of clusters should be updated too with the lifetime of the network. So authors have suggested that the dynamic optimum number of cluster base on the real time of sensor nodes in the whole network, and it brings practical optimized effect in WSNs. Yang et al. [42], reduced the energy consumption of the network and tried to avoid the severe synchronization needs of Time Division Multiple Access. This algorithm is first accessible for small networks, under the hypothesis of identical expected distance of all CHs from the BS. Then it is extended for large networks to consider the case when the distances of various sections of the network from the BS may be different. This new architecture is easily implementable, scalable, and reduces the hardware complexity of the sensor nodes.

Chan et al. [48] proposed a Fixed Optimal Cluster (FOC) numbers that is to examine the entire network. There are two different types of the optimal cluster numbers depending on the position of the base station. One is that the base station is set up far farther away from the sensing field, and another is that the base station is set up at the center of the sensing field Results show that network lifetime enhanced very well. Shunjie et al. [52] proposed new model for finding the optimal number of CHs which is based on the LEACH energy dissipation model. Optimal number of clusters depends upon the number of frames in steady-state phase and the distance between CHs and the BS.

Yang et al. [61] have proposed a more reasonable energy consumption model called Optimal Energy Consumption Model (OECM) in a homogeneous network. It shows that the optimal number of cluster heads not only depend on node density, but also on the size of sensing field, circuit energy dissipation, packet length. Navid et al [62], analytically provides the optimal number of clusters that minimizes the total energy expenses in the networks, where all sensor nodes communicate data through their elected CHs to the BS in a distributed fashion. The results show that: 1 Under certain condition, the optimal number of clusters can be independent of the size of sensing field 2: The energy consumption of the transmitter circuitry has no impact on the optimal number of clusters, and 3: The energy consumption of the receiver electronics can substantially change the optimal number of clusters and more importantly; it can decide on whether or not is it worth performing clustering. Li et al. [55], have proposed, an uneven virtual grid-based clustering routing protocol. It works in concentric circular forms through the base station-assisted positioning. It's providing better energy efficiency and balanced network load.

Xia et al. in [41], proposed Local Negotiated Clustering Algorithm (LNCA) that tries to minimize the overall energy cost of the network by using the similarity of nodes. Experimental results of the algorithm

indicate that 2- hop LNCA outperforms existing LEACH [31] algorithms in terms of energy consumption. Optimal number of clusters depends on the sensing field area, node density in the network the length of data, the rate at which data is generated at each node and the length of time for inter cluster communication. Halgamuge et al. in [51] have developed an energy model and used this model to assess energy expenditure and node lifetime for a sensor network with fixed configuration. The authors also have two observations 1: The optimal number of clusters increases with the increase of free space fading energy 2: The number of clusters does not play a significant role for realistic size sensor networks if the free space fading energy is low. For large networks, cluster optimization is still important, even if free space fading energy is low. Li et al. [58] presented an analytical model for finding the optimal number of clusters, for minimizing the communication costs in a clustered sensor network. In this, sensor nodes are placed in the sensing field in a random and distributed manner according to a homogeneous poisson point process.

Tandon in [64], focused on the analysis of extravagant energy consumption within a uniform CH election model and suggested a model to reduce the overall consumption of energy usage amongst the CHs in the WSNs. The optimal number of clusters depends on the distance from the BS. Chen et al.[63] proposed CH optimization based on energy. In this, the authors considered a threshold value and the residual energy of nodes, to optimize the selection of a cluster head. Results show this algorithm can prolong the network lifetime efficiently compared with LEACH protocol. The algorithms based on uniform node distribution have following problems:

- ✓ Cluster head selection is uncontrolled.
- ✓ Uniform sensor node distribution.
- ✓ Cannot be applied to all the practical cases.

4.1.2 Probabilistic and Non-uniform Node Distribution

Tripathi et al. [66] introduced clustering of non-uniform random distributed nodes, and calculated the optimal number of clusters in WSN. Results show that there is balanced energy expenditure in this non-uniform clustering. Dabirmoghaddam et al. [67], that the general problem of optimal clustering is NP-hard. They try to optimize the algorithm and produce the best possible clustering of the network in terms of energy consumption. It is found that with non-uniform clustering in heterogeneous WSNs, clusters are more energy-efficient in WSNs with spatial data correlation.

4.1.3 Non-probabilistic Algorithms

In this type of clustering algorithms, selection of cluster heads and cluster formation is based on deterministic criteria such as connectivity, degree and information received from the other closely located group.

Pattem et al. 36] found the optimal number of clusters, which depend upon the distance from the cluster head to the BS and the degree of correlation. As we know that the small cluster sizes and large cluster sizes perform well at low and high correlations respectively. Nevertheless, it appears that an intermediary cluster size performs correctly across the range of correlation values that is called “near-optimal” static cluster size. The value of near optimal cluster size depends only on BS position and the number of nodes. Chen et al. [49] worked on a multi-cluster sensor network, which is used for source extraction in a sensing field. Performance of source extraction and the complete energy consumption in

the sensor network depends upon the number of clusters. Performance of source extraction and the entire energy consumption in the sensor network can be affected by different factors: unreliable wireless links, sensor failure, etc.

Raghuvanshi et al. [50] have used an objective function based method to divide a data set into a set of clusters. In comparison to standard clustering, fuzzy clustering offer to assign a data point to more than one cluster, so that overlapping clusters can be handled confidently. In this paper Gustafson- Kessel (GK) algorithm is used for partitioning the sensor nodes into clusters and optimal number of clusters are determined using Xie-Beni validation index. This algorithm uses fuzzy clustering for partitioning the data into a pre-defined number of clusters with fuzzy boundaries. G-K clustering algorithms can find clusters of approximately equal areas, thus load balancing is good. In a G-K clustering, number of clusters is a range of numbers rather than a unique number. G-K clustering provides more coverage range comparisons to LEACH clustering algorithm. Selvakennedy et al. [46] tried to find the appropriate number of clusters with well-balanced memberships. The proposed algorithm is speedy with very partial overhead. Due to the robustness of any biologically-inspired algorithm, this protocol could handle unexpected circumstances in the environment and node failures.

Wang et al. [53] proposed the main challenges in deploying a high dense cluster based sensor network. In this paper, the authors have considered a basic Observational Area (OA) in WSNs and determined optimal numbers of clusters in basic OAs from the cross layer approach. The authors find a number of clusters from different layer aspects, and they have shown and concluded that from the physical layer point of view if there are more clusters results in more energy savings. From MAC layer and network layer point of view, fewer cluster results in more energy savings. Hence, authors have considered and developed a cross layer optimized model for physical, MAC, and Network Layer (PHY/MAC/NET). They have shown that optimizing the number of clusters in a sensor network becomes a cross layer tradeoff design issue among the required transmission power in the physical layer, cluster representative in the MAC and hop count in the network layer.

Wang et al. [56] proposed fuzzy based clustering, as we know that Fuzzy c-Means (FCM) and its derivatives suffer from two problems: local minima and cluster validity—which have a straight impact on the formation of the final clustering. The problem of local minima can be short out by optimization and center initialization strategies. This paper proposes a center initialization approach based on a minimum spanning tree to keep FCM from local minima. Raghuvansi et al. [60], describes a method for finding a fuzzy membership matrix that provides cluster membership values for all the objects based on the proximity matrix. Membership matrix related with fuzzy is found by first finding a set of vectors that approximately have the same inter-vector euclidean distances as the proximate that are provided. A dimension of these vectors can be very low (less or equal to 5). FCM algorithm is used for the optimal number of cluster calculation in WSNs. The authors found that this method to be very effective and no more computationally expensive than other relational data clustering methods. The FCM algorithm is more energy efficient compared to a G-K clustering algorithm. There are several validity measures proposed in the literature, given the optimal number of clusters following seven measures are considered. 1: Partition Coefficient (PC) (L), 2: Xie and Beni's Index (XB) (L), 3: Classification Entropy (CE) (M), 4: Partition Index (SC) (L), 5: Separation Index (S) (L), 6: Dunn's Index (DI) (M), 7: Alternative Dunn Index (ADI) (L).

4.2 Heterogeneous sensor networks

There are three common types of hardware heterogeneities in WSNs: **Computational heterogeneity** in which the heterogeneous node has a more powerful microprocessor, and more memory, than the normal node. **Link heterogeneity:** In which some nodes have long-distance highly reliable communication links (IEEE 802.11 connectivity) than a normal node. **Energy heterogeneity** where nodes have unlimited energy resources or battery is replaceable. The most important heterogeneity is the energy heterogeneity because both computational heterogeneity and link heterogeneity will consume more energy resources. Heterogeneity provides improvement in terms of response time, reliability and lifetime [68, 69]. If there is no energy heterogeneity, computational and link heterogeneity will impact the whole sensor network negatively, resulting decrement of the network lifetime [28, 30, 47].

Mhatre et al. [34] determined expressions for the required number of CH along with type of communication like single hopping, multi-hopping and hybrid modes. Authors have two types of sensor nodes, which are less robust due to heterogeneity of the nodes. If any cluster head (CH) nodes fail, the system no longer will function. In LEACH protocol, the system is more robust because every node is competent of acting as a CH, and hence the subside of a few nodes does not seriously affect the working of the whole system. Smaragdakis et al. [35] proposed, a heterogeneous-aware protocol; it had applications where the feedback from the sensor network must be reliable.

Kumar et al. [47] have introduced hierarchically clustered network in which sensor nodes have energy heterogeneity, means some of the nodes equipped with additional energy. Kumar et al. [54] have developed an energy-efficient cluster head election (EECHE) protocol for heterogeneous WSN with three types of sensor nodes. Some of the sensor nodes are equipped with the extra energy resources than the other nodes. Results show that EECHE can enhance the lifetime and stability of the system and provide better results than LEACH protocol. Tuah et al. in [65] proposed heterogeneous sensor networks and investigated the optimal number of clusters, which can minimize the energy consumption per round. In this two type of radio energy models are used, in which intra cluster communication using free space model and inter cluster communication using multipath model respectively.

It is seen that the performance of clustered WSNs is not always better than the performance of non-clustered WSNs under certain condition. The ability of clustered WSNs to outperform their non-clustered counterparts depends on number of nodes, the total number of cluster heads, and average factor of in-cluster data reduction [10].

Table 3: Symbols and their meaning used in different algorithms surveyed

Symbol	Abbreviation	Symbol	Abbreviation
K_{opt}	Optimal number of clusters or CHs	N	Number of sensor nodes
M	Side of the Square	d_{toBS}	Distance between the cluster head and base station
$E_{elec} = E_{tx\_elec}$ $=E_{rx\_elec}$	Energy dissipated to run the transmitter or the receiver circuitry to transmit or receive one bit of the data packet	E_{DA}	Data Aggregation Energy
ε_{fs}	It is energy dissipation of the transmission amplifier to convey one bit of data packet to the receiver node with a distance of d=1m away, for n=2 (Free Space)	ε_{mp} =E_{amp}	It is energy dissipation of the transmission amplifier to convey one bit of data packet to the receiver node with a distance of d=1m away, for n=4(Multipath)
d	Distance between transmitter and receiver	d_{toCH}	Distance between sensor nodes and cluster head
m	Data compression rate	B	Distance from the centre of sensing area to the outside location of the base station.
i^{nth}	i^{th} *ring*	A_{net}	Area of circular field
R_{net}	Radius of circular fields	λ_i	Variable density parameter
j	No. of annular bands of same radial width	N_i	No. of nodes in an annular band of $\lambda_i \times A_i$
L	Signal frame with L-samples	$E\left\{ d_{toCH}^4 \right\}$	Expected Value of distance between cluster heads and base station
λ	Intensity of homogeneous spatial Poisson process	$l^1 + \mu^1 d^{k1}$	Energy Spent in transmitting a packet from cluster head to Base Station
α_i	Hardware cost of the node	β	Used to model node battery
T	Data gathering cycle	μ	Propagation Loss Term
n	Propagation loss exponent	r	Radio range of node
c	Average of C_i	c_i	Degree of correlation between a sensor and its neighborhood sensors
f	Number of data frames	l_1	Bit length of control information
l_2	Bit length of data information	T	The time to transmit one byte on the network
T_{inter}	time devoted for inter cluster communication	k_{inter}	$\frac{T_{inter}}{T}$
$d_{broadcast}$	Distance between the CH and the farthest point of observed area	c_o	Spatial Correlation
D	No. of hop between CHs and BS	ρ	Energy Multiplication factor of the super advanced node
γ	Data Fusion Rate	α	Energy Multiplication factor of the advanced node
m_0	Percentage of Advanced Node	m_1	Percentage of super Advanced Node
η	Required received power	L_0	Path loss at distance k_o
P_e	Receiver electronics energy	d_{OA}	Minimal distance that results in uncorrelated information
R	Radius of circular Field	***k***	Propagation loss constant for

			communication between the cluster head
κ^1	Propagation loss constant for communication between the cluster heads and Base Station	z	Number of hops between node and Base station
n_o	Type "0" Nodes	d^o	Discrete time
μA^K	Amount of Energy Spent in the RF Amplifier to counter propogation Loss	δ	Distance between two rings
d_{toCh1}	Average Distance between two CHs	N_1	Number of nodes in first ring
E_{tranCH}	Energy Dissipation due to operating mode at the CH per Round	E_{loggCH}	Energy consumed for logging sensor reading at the CH per round
E_{SensCH}	Total Energy dissipation for sensing activity at the CH per round	L	Position of BS from the centre of sensing field
E_c	Energy Consumed in computation of Data		

Table 4: Comparative Analysis of algorithms based on network type, Analytical value of optimal number of clusters and Base Station Position (network topology)

Paper Ref.	Network Type	The analytical value of K_{opt}	Base Station Position
[31]	Homogenous	$K_{opt} = \left(\frac{N}{2\pi} \times \frac{\varepsilon_{fs}}{\varepsilon_{mp}}\right)^{\frac{1}{2}} \frac{M}{d^2_{toBS}}$	Center of Sensing Field
[32]	Homogenous	$K_{opt} = \left\{\frac{N}{2} \times \left(\frac{M^2}{2\pi}\right)^{\frac{n}{2}} \times \frac{1}{(d_{broad\,cost})^n}\right\}^{\frac{n}{2}+1}$	Outside of sensing field
[34]	Heterogeneous	$K_{opt} = \left\{\frac{kn_o\beta T\mu A^k}{2(\alpha_1 + c\beta T(l^1 + \mu^1 d^{k^1}))}\right\}^{\frac{2}{k+2}}$	Outside of sensing field (Single Hopping)
		$K_{opt} = \left\{\frac{n_o\beta T r_H{}^2(2l + \mu r^k)}{r^2(\alpha_1 + c\beta T(l^1 + \mu^1 d^{k^1}))}\right\}^{\frac{1}{2}}$	Outside of sensing field (Multi Hopping)
[35]	Heterogeneous	$K_{opt} = \sqrt{\frac{N}{2\pi}} \times \frac{2}{0.765}$	Center of Sensing Field
[36]	Homogenous	$K_{opt} = \sqrt{2Dc_o}$	Moving
[37]	Homogenous	$K_{opt} = \left\{\frac{0.5855}{\varepsilon_{mp}} \times \frac{N \times \varepsilon_{fs} \times M^2}{d^4_{toBS} - E_{elec}}\right\}^{\frac{1}{2}}$	Outside of sensing field
[38]	Homogenous	$K_{opt,1} = M.\left\{\frac{N\varepsilon_{fs}}{\pi(\varepsilon_{mp} d^4_{toBS} - E_{elec})}\right\}^{\frac{1}{2}}$	Outside of sensing field
[39]	Homogenous	$K_{opt} = \left\{\frac{cN}{2(1-c)}\right\}^{\frac{2}{3}}$	Center of Sensing Field
[40]	Homogenous	$K_{opt} = \sqrt{4Z - 1}$	Center of Sensing Field
[41]	Homogenous	$K_{opt} = \frac{1}{36}(1 + P + P^{-1})^2$ $P = (1 + 54N + 6 \times \sqrt{3 \times N + 81N^2})^{\frac{1}{3}}$	Center of Sensing Field
[42]	Homogenous	$K_{opt} = \left\{\frac{N\varepsilon_{amp} M^2}{3E_{elec} k_{inter} + 2\varepsilon_{amp} M^2 K_{data}}\right\}^{\frac{1}{2}}$	Center of Sensing Field
[43]	Homogenous	$K_{opt} = \left(\frac{N}{2\pi} \times \frac{\varepsilon_{fs}}{\varepsilon_{mp}}\right)^{\frac{1}{2}} \frac{M}{d^2_{toBS}}$	Inside Sensing field
[46]	Homogenous	$P_{opt} = \frac{1}{2} \times \sqrt{\frac{\varepsilon_{fs}}{\lambda[\frac{\sqrt{\frac{1}{6M}}}{r}(2E_{elec} + \varepsilon_{fs} r^2) - 2E_{elec}]}}$	Center of Sensing Field

[47]	Heterogeneous	$K_{opt} = \left(\frac{N}{2\pi} \times \frac{\varepsilon_{fs}}{\varepsilon_{mp}}\right)^{\frac{1}{2}} \frac{M}{d^2_{toBS}}$	Center of Sensing Field
[48]	Homogeneous	$K_{opt} = \sqrt{N}$	Center of Sensing Field
		$K_{opt} = \sqrt{6N} \times \sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{mp}}} \times \frac{M^2}{M^2 + 6B^2}$ $B \geq \sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{mp}} - \frac{M^2}{6}}$	***When*** $d_{toBS} \geq \sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{mp}}}$ For multipath fading model
		$K_{opt} = \sqrt{\frac{N\varepsilon_{fs}}{6\varepsilon_{mp}}} \times \frac{M}{d^2_{toBS}}$	$d_{toBS} \leq \sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{mp}}}$ ***For free Space Model fading model***
[49]	Heterogeneous	$K_{opt} = \left\{\frac{\varepsilon_{fs} d^0 lNM^2 + 4\pi d^0 N^2 E_c}{2\pi[(8-2l)d^0 E_{elec} + 8\varepsilon_{mp} d^0 E\{d^4_{toCH}\} + (2l+7)d^0 E_c]}\right\}^{\frac{1}{2}}$	Above the center of field
[50]	Homogeneous	Numerical results is given	Center of Sensing Field
[51]	Homogeneous	$K_{opt} = \sqrt{\frac{N_s}{6}} \times \frac{M}{d^2_{toBS}} \times \sqrt{\frac{E_{fs}}{D_\alpha}}$ $D_\alpha = (E_{amp} + E_{sensCH} + E_{tranCH} + E_{\log gCH})$	Out-side of field
[52]	Homogeneous	$K_{opt} = \left\{\frac{(l_1 + l_2 f) N\varepsilon_{fs} M^2}{2\pi[\varepsilon_{mp}(l_1 d^4 + l_2 f d^4_{toBS}) - (2l_1 + l_2 f)E_{elec}]}\right\}^{\frac{1}{2}}$	Center of Sensing Field
[53]	Homogeneous	$K_{opt} = \begin{cases} \left(\frac{\left(\frac{n}{2}-1\right)\xi}{2P_e}\right)^{\frac{n}{2}} & for\ n > 2 \\ 1 & for\ n = 2 \end{cases}$ $\xi = \eta L_o \left(\frac{\sqrt{5} \times d_{OA}}{k_o}\right)^n$	Center of Sensing Field (Physical and MAC)
		$K_{opt} = \begin{cases} \left(\frac{z}{2p_e}\right)^{\frac{2}{n}} & for\ n > 2 \quad z = \left(\left(\frac{n}{2}-1\right)\right) \times \xi \times 10^{Q^{-1}(\Theta)\frac{\sigma}{10}} \\ 1 & for\ n = 2 \end{cases}$	Center of Sensing Field (Shadowing effects)
[54]	Heterogeneous	$K_{opt} = \left(\frac{N}{2\pi} \times \frac{\varepsilon_{fs}}{\varepsilon_{mp}}\right)^{\frac{1}{2}} \frac{M}{d^2_{toBS}}$	Center of Sensing Field
[55]	Homogeneous	$K_{opt} = \sqrt{\frac{N_1}{2}} \times \frac{\delta}{\sqrt{E[d^2_{toCH1}]}}$	Center of Sensing Field
[56]	Homogeneous	Referred [56]	N/A
[57]	Homogeneous	based on Experimentally Analysis	N/A
[58]	Homogeneous	$K_{opt} = \left\{\frac{3N^2\varepsilon_{fs}}{\pi\lambda(2M^2\varepsilon_{fs} - 3E_{elec})}\right\}^{\frac{1}{2}}$	The BS is located in one vertex of an square area
[60]	Homogeneous	Numerical Results given	Center of Sensing Field
[61]	Homogeneous	Numerical Results given	Center of Sensing Field

[62]	Homogeneous	$K_{opt} = \sqrt{\frac{N\varepsilon_{fs}A}{2\pi(\varepsilon_{amp}d_{toBS}^{n} - E_{rx-elec})}}$	Dynamic base Station
[63]	Homogeneous	$K_{opt} = \left\{\frac{NM^2\varepsilon_{fs}}{0.342M^4\varepsilon_{amp} - E_{elec}}\right\}^{\frac{1}{2}}$	Center of Sensing Field
[64]	Homogeneous	$K_{opt} = \left\{\frac{3\varepsilon_{fs}N(2i+1)^3}{2\varepsilon_{mp}R_{net}^2((i+1)^6 - i^6)}\right\}^{\frac{1}{2}}$	Center of Sensing Field
[65]	Heterogeneous	$K_{opt} = \frac{M\sqrt{3N}}{\sqrt{2\pi}}\left\{\frac{\varepsilon_{fs}\{(1-(m_0+m_1))+\alpha m_0+\rho m_1\}}{E_{elec}\left\{\left(\frac{3}{\gamma} - N(1+\alpha+\rho) + \frac{3\varepsilon_{mp}d_{toBS}^4}{\gamma}\right)\right\}}\right\}$	Center of Sensing Field
[66]	Homogeneous	$K_{opt} = \frac{N^{\frac{3}{2}} \times R}{\sqrt{2} \times \sum_{i=1}^{j}\lambda_i A_i R_i}$	Center of Sensing Field
[67]	Homogeneous	Numerical Value is given	Center of Sensing Field

Table 5: This table is well taken from [62] for comparative Analysis of Algorithm based on sensing field, radio model, base station position and Analytical value of optimal number of clusters

Reference [62]				
Sensing field	**Radio model**	**Location of BS**	**Optimal number of clusters**	**Optimal number of clusters (Small E_{Rx})**
Square $(M \times M)$	**Free space**	**Center**	$K_{opt} = \sqrt{\frac{N\varepsilon_{fs}M^2}{2\pi(\varepsilon_{fs}\frac{M^2}{c} - E_{rx\_elec})}}$	$K_{opt} = \sqrt{\frac{3N}{\pi}} = 0.977\sqrt{N}$
		Corner	$K_{opt} = \sqrt{\frac{N\varepsilon_{fs}M^2}{2\pi(\varepsilon_{fs}\frac{2M^2}{2} - E_{rx\_elec})}}$	$K_{opt} = \sqrt{\frac{3N}{4\pi}} = 0.489\sqrt{N}$
		Side midpoint	$K_{opt} = \sqrt{\frac{N\varepsilon_{fs}M^2}{2\pi(\varepsilon_{fs}\frac{5M^2}{12} - E_{rx\_elec})}}$	$K_{opt} = \sqrt{\frac{6N}{5\pi}} = 0.618\sqrt{N}$
		Outside on the (Axis of symmetry)	$K_{opt} = \sqrt{\frac{N\varepsilon_{fs}M^2}{2\pi(\varepsilon_{fs}(\frac{M^2}{6} + L^2) - E_{rx\_elec})}}$	$K_{opt} = \sqrt{\frac{NM^2}{2\pi(\frac{M^2}{6} + L^2)}}$
	Two ray	**Center**	$K_{opt} = \sqrt{\frac{N\varepsilon_{fs}M^2}{2\pi(\varepsilon_{mp}\frac{7M^2}{180} - E_{rx\_elec})}}$	$K_{opt} = \sqrt{\frac{90N\varepsilon_{fs}}{7\pi\varepsilon_{mp}M^2}}$
		Corner	$K_{opt} = \sqrt{\frac{N\varepsilon_{fs}M^2}{2\pi(\varepsilon_{mp}\frac{28M^4}{45} - E_{rx\_elec})}}$	$K_{opt} = \sqrt{\frac{45N\varepsilon_{fs}}{56\pi\varepsilon_{mp}M^2}}$
		Side mid point	$K_{opt} = \sqrt{\frac{N\varepsilon_{fs}M^2}{2\pi(\varepsilon_{mp}\frac{193M^4}{720} - E_{rx\_elec})}}$	$K_{opt} = \sqrt{\frac{360N\varepsilon_{fs}}{193\pi\varepsilon_{mp}M^2}}$
		Out (on the axis of symmetry)	$K_{opt} = \sqrt{\frac{N\varepsilon_{fs}M^2}{2\pi(\varepsilon_{mp}(\frac{7M^4}{180} + \frac{2}{3}M^2L^2 + L^4) - E_{rx\_elec})}}$	$K_{opt} = \sqrt{\frac{N\varepsilon_{fs}M^2}{2\pi\varepsilon_{mp}(\frac{7M^4}{180} + \frac{2}{3}M^2L^2 + L^4)}}$
Circle (R)	**Free space**	**Center**	$K_{opt} = \sqrt{\frac{N\varepsilon_{fs}R^2}{R^2\varepsilon_{fs} - 2E_{rx\_elec})}}$	$K_{opt} = \sqrt{N}$
		Circumference	$K_{opt} = \sqrt{\frac{N\varepsilon_{fs}R^2}{3R^2\varepsilon_{fs} - 2E_{rx\_elec})}}$	$K_{opt} = 0.577\sqrt{N}$
		Outside	$K_{opt} = \sqrt{\frac{N\varepsilon_{fs}R^2}{2(\varepsilon_{fs}(\frac{R^2}{2} + L^2) - E_{rx\_elec})}}$	$K_{opt} = \sqrt{\frac{NR^2}{2(\frac{R^2}{2} + L^2)}}$

	Two ray	**Center**	$K_{opt} = \sqrt{\frac{N\varepsilon_{fs}R^2}{2(\varepsilon_{mp}\frac{R^4}{3} - E_{rx\_elec})}}$	$K_{opt} = \sqrt{\frac{3N\varepsilon_{fs}}{2\varepsilon_{mp}R^2}}$
		Circumference	$K_{opt} = \sqrt{\frac{N\varepsilon_{fs}R^2}{2(\varepsilon_{mp}\frac{10R^4}{3} - E_{rx\_elec})}}$	$K_{opt} = \sqrt{\frac{3N\varepsilon_{fs}}{20\varepsilon_{mp}R^2}}$
		Outside	$K_{opt} = \sqrt{\frac{N\varepsilon_{fs}R^2}{2(\varepsilon_{mp}(\frac{R^4}{2} + L^4 + 2R^2L^2) - E_{rx\_elec})}}$	$K_{opt} = \sqrt{\frac{N\varepsilon_{fs}R^2}{2\varepsilon_{mp}(\frac{R^4}{3} + L^4 + 2R^2L^2)}}$

Table 6: Time-line comparative analysis of algorithms based on cluster variability, year of publication, node distribution, energy model and type of sensing fields

Reference	Year of Publication	Cluster Variability	Node Distribution	Energy Model Used	Sensing Fields
[31]	2002	Fixed	Uniform and random	[31]	Square
[32]	2003	Adaptive	Uniform and random	[63]	Square
[34]	2003	Adaptive	Uniform and random	[31]	Circular
[35]	2004	Adaptive	Uniform and random	[31]	Square
[36]	2004	Fixed	Uniform and random	N/A	Square
[37]	2005	Fixed	Uniform and random	[31]	Square
[38]	2006	Adaptive	Spatial Poisson	[31]	Square
[39]	2006	Adaptive	Uniform and random	[31]	N/A
[40]	2006	Adaptive	Linearly placed nodes	N/A	N/A
[41]	2007	Adaptive	Uniform and random	N/A	Square
[42]	2007	Fixed	Uniform and random	[31]	Square
[43]	2007	Adaptive	Uniform and random	[31]	Square
[46]	2007	Adaptive	Spatial Poisson	N/A	Circular
[47]	2008	Adaptive	Uniform and random	[31]	Square
[48]	2008	Fixed	Uniform and random	[31]	Square
[49]	2009	Adaptive	Uniform and random	[31]	Square
[50]	2009	Adaptive	Uniform and random	N/A	Square
[51]	2009	Adaptive	Uniform and random	[51]	Square
[52]	2009	Adaptive	Uniform and random	[31]	Square
[53]	2009	Adaptive	Uniform Distribution[75]	[31] &[71]	Square
[54]	2009	Adaptive	Uniform and random	[31]	Square
[55]	2009	Adaptive	Uniform and random	[31]	Concentric Ring
[56]	2009	Adaptive	……	N/A	N/A
[57]	2010	Adaptive	Uniform and random	No Specific Energy Model	Square
[58]	2010	Adaptive	Spatial Poisson	[58]	Square
[60]	2010	Adaptive	Uniform and random	N/A	Square
[61]	2010	Adaptive	Spatial Poisson	[35]	Square
[62]	2011	Adaptive	Uniform and random	[31]	Circular and square
[63]	2011	Adaptive	Spatial Poisson	[74]	Square
[64]	2012	Adaptive	Uniform and random	[31]	Circular
[65]	2012	Adaptive	Uniform and random	[31]	Square
[66]	2013	Adaptive	Spatial Poisson	[31]	Circular
[67]	2014	Adaptive and static	Uniform and random	[31]	Circular

Table 7: Comparative Analysis of algorithms based on their strengths, weaknesses and applications

Refer.	Strengths	Weaknesses	Applications
[31]	It reduces data collisions. It is more robust. The random mechanism of cluster head selection has many advantages such as easy realization, easy operation, and better scalability.	The Energy model used is too ideal to some extent, especially when the scale of sensing field is large. Overhead generated is very high.	This algorithm provides the high performance needed under the tight constraints of the wireless channel.
[32]	This algorithm outperform LEACH-A in a large class of situations and applications and especially when the final receiver is closer to the sensors and the deterministic attenuation due to the path-loss has a value of *no* larger than 2.5.	Aggregation of data is not used in this algorithm	It can used for the monitoring of a big car-park in which sensors, distributed in the area, interact to communicate with an external receiver (mounted over a car) the best way to reach the closest free place.
[34]	It's required less hardware and software complexity	Less robust	It is applicable to the overall design problem through a data aggregation model
[35]	SEP yields longer stability region for higher values of extra energy brought by more powerful nodes.	Due to practical/cost constraints, it is not always possible to satisfy the constraints for optimal distribution between different types of nodes.	One of the applications could be the re-energization of sensor networks. There are also applications where the spatial density of sensors is a constraint.
[36]	It reduces the overhead generated	This algorithm has ignored lossy compression	This algorithm concentrates on applications which involve continuous data gathering for large scale and distributed physical phenomena using a dense WSNs.
[37]	Its minimizing the energy consumption of the system	This method cannot be applied in real time based model	This can be applied to the fact that the optimal number of clusters-heads determines as the different density of sensors in a bounded area
[38]	Energy consumption of the network decreases as the number of levels is increased	Increased overhead and delay relative to single level clustering	It is expected to be most suited for delay insensitive applications where data aggregation is required
[39]	It provides the efficient utilization of data correlation.	Type of Routing used, allows for maximum possible aggregation at each hop, which may not always be practically possible to implement.	It is applicable for the place where energy consumption required with respect to data correlation of the nodes
[40]	Provides Energy conservation in the networks	It is focusing on the optimal cluster size, is based on the assumption that each issue cluster area spans over one, or more, square shaped clusters, which may not always be true in real world WSN environments.	We can take a decision whether clustering will be useful or not.
[41]	Intra-cluster communication can be completely avoided by this algorithm which is not true in the case of LEACH protocol.	It does not include the development of cost-effective and rapid re-clustering mechanisms according to the environmental changes.	It is highly effective in minimizing in-network data-reporting traffic and, accordingly, in reducing the energy usage of individual sensor nodes
[42]	It reduces the overhead of the system	Complexity of the algorithms is high	It can be used for dynamic network

[43]	Network lifetime is enhanced	System complexity is increased	It is more acceptable for monitoring and measuring larger scale of water environment than before [44-45].
[46]	This clustering protocol exhibits fault-tolerance in the case of node failures compared deterministic clustering approaches in the literature.	Identification of the optimal number of ants is a little bit tough process.	Due to the robustness of any biologically-inspired algorithm, this protocol could handle unexpected circumstances in the environment and node failures. This approach may also be useful in applications that require an in-network actuation, to assist in the sensor–actuator coordination.
[47]	It has extended the lifetime of the network by 10% as compared with LEACH in the presence of the same setting of powerful nodes in a network	This algorithm is valid only when intra-cluster communication phase is long enough	This algorithm can be used for monitoring the application
[48]	Its reduces data collisions It is more robust	Overhead generated is very high	It provides best performance needed under the rigid constraints of the wireless channel.
[49]	Performance is greatly improved by adopting a multi-cluster structure	The multi-hop setup was not explored in this algorithm.	This paper studies a multi-cluster sensor network, which is applied for source extraction in a sensing field
[51]	It has a new, realistic and comprehensive energy model for wireless sensor networks	…….	With the help this algorithm we can take decision whether we should do optimal clustering or not the basis of free space fading energy
[52]	It has better performance than LEACH	This algorithm is not applicable to the real time scenario. It is not applicable for asymmetrical channel	It is applicable for dynamic topology
[53]	This algorithm provides an optimal cluster number that can still effectively function, regardless of the different densities of sensors	The complexity of this algorithm is high	The proposed cross-layer analytical model can facilitate the design of the optimal number of clusters for a sensor network in different radio environments
[54]	It has better performance than LEACH	Overhead generated is very high Fault tolerance is also a major issue	This algorithm can equally apply to small sized wireless networks.
[55]	Network lifetime increase.	It's not providing good connectivity and coverage	This algorithm has application where we location based application
[56]	It is better than or comparable with CCIA and Kd-tree [72,73] in terms of pattern recognition rate.	This algorithm is computationally expensive in some cases	It can apply even though the structure of the dataset is complex
[57]	It shows that 1-hop clustering performs best for a large spectrum of different network sizes, node densities and the number of base stations.	……..	It can be used for real transmission
[58]	Les overhead	Algorithms only considered error free communication and do not account for the channel contention	It is applicable for determining the optimal number of clusters based on other MAC and routing algorithms.
[60]	More Stable region.	……..	It can be used for approximation problems and recognition of geometrical shapes in image processing
[61]	Provide Scalable Network	This algorithm does not use a real scenario energy model	Applicable for different network environment and parameters

[62]	Enhanced Network Lifetime	Energy wastage due to variability of the number of clusters	This is applicable for providing the analytical model for all types of sensing fields and the position of the base station in the scenario
[63]	This algorithm can prolong the network lifetime efficiently compared with LEACH protocol	It cannot ensure uniform distributed of cluster heads in the network for it does not consider the location of nodes.	It is useful for homogeneous type of applications
[64]	It provide on an average a 28 % reduction in total energy usage over other existing algorithm	It can not apply for location aware applications	This scheme would be very suitable in applications where either the required accuracy in data is low or the data has very high redundancy
[65]	Its provide the perfect selection of optimal number of clusters	This algorithm has less fault tolerance	The application of this technique in WSNs is possible due to the redundancy in sensor readings and the large number of nodes deployed
[66]	Energy Efficient	Complexity of the system increases	It can be applied to practical scenario like a battlefield where we are placing nodes in an uncontrolled manner
[67]	High Network lifetime	……	It is applicable for non-uniform distributed nodes in WSNs

5 Open Issues and Challenges

The open issues and challenges in WSNs, which can serve as research topics for future work are summarized below

- ✓ Finding optimal number of clusters with moving nodes that provides good coverage and energy-efficient system.
- ✓ Finding optimal number of clusters for non-uniform node distribution (Gaussian distribution) in application for remote places, where sensor nodes are dropped by helicopter.
- ✓ How to select optimal number of clusters without knowing the location of the nodes?
- ✓ Optimal number of clusters using a cross layer approach along with random traffic and data aggregation model.
- ✓ Finding optimal number of clusters using more practical energy model other than an existing one [31]
- ✓ Calculating optimal number of CHs by using probabilistic sensing model of the nodes (Elfes Sensing model)

6 Conclusion

WSNs are an emerging technology that has been attracting large pool of researchers in recent years. Optimal number of clusters plays a crucial role in the performance of WSNs, in terms of system scalability, energy efficiency, collision reduction, network lifetime, latency, and efficient routing backbone in the network. We have surveyed the state-of-art of different algorithms for finding the optimal number of clusters that have been analytically reported in the literature of WSNs, and have presented the methodologies utilized by different authors to calculate the optimal number of clusters in WSNs from an energy-efficiency point of view with the help of taxonomy. This paper discusses the fundamental concepts of clustering, need of clustering, optimal clustering and need of optimal clustering in WSNs. We have compared the optimal clustering algorithms based on cluster variability, heterogeneity, analytical formula for optimal number of clusters, type of energy model used, type of node distribution, type of sensing field, the position of the base station, strengths, weaknesses and

applications of each and every algorithm for calculating an optimal number of clusters. We have also discussed the impact of different levels at which WSNs is modeled viz. Radio Energy Model Level, Network Model Level and Clustering Level for selecting optimal number of clusters along with open issues and challenges in WSNs.

REFERENCES

[1]. Pottie, G.; Kaiser, W. Wireless Integrated Network Sensors. *ACM Communications, 2000, 43, 5,p. 51–58.*

[2]. Kushwaha, S.; Kumar, V.; Jain, S. Node Architectures and Its Deployment in Wireless Sensor Networks: A Survey. In *High Performance Architecture and Grid Computing, published Springer Berlin Heidelberg, 2011, p. 515-526*.

[3]. Yu, J. Y.; Chong, P. H. J. A Survey of Clustering Schemes for mobile ad hoc networks. *IEEE Communications Surveys Tutorials, 2005, 7*,p. *32–48*.

[4]. Kumar, V.; Jain, S.; Tiwari, S. Energy Efficient Clustering Algorithms in Wireless Sensor Networks: A survey. *IJCSI International Journal of Computer Science Issues, 2011, 8(5),p. 1694-0814.*

[5]. Kumar, V.; Tiwari, S. Energy Efficient Mechanisms in Wireless Sensor Networks: A survey. *International Journal of Advanced Research in Computer Science, 2011, 2(5),p. 595-604.*

[6]. Abbasi, A. A.; Younis, M. A Survey on Clustering Algorithms for Wireless Sensor Networks. *Elsevier Science direct Computer Communications, 2007, 30,p. 2826–2841.*

[7]. Deosarkar, B. P.; Yadav, N. S.; Yadav, R. Cluster head selection in clustering algorithms for wireless sensor networks: A Survey. *Proceedings of International Conference on Computing, Communication and Networking, 2008, p. 1–8.*

[8]. Akyildiz, I. F.; Vuran, M. C.Wireless Sensor Networks. *A John Wiley and Sons, Ltd, Publication, 2010.*

[9]. Zhang, Y.; Yang, L.T.; Chen J. RFID and Sensor networks: Architectures, Protocols, Security, and Integrations. *CRC press Taylor and Francis 2010*.

[10]. Heinzelman, W.R.; Chandrakasan, A.; Balakrishnan, H. Energy-efficient communication protocol for wireless microsensor networks. *Proceeding 33rd Hawaii International Conference on System Sciences, 2000, p. 3005–3014.*

[11]. Dali, W.; Chan, H. A survey on cluster schemes in ad hoc wireless networks. *Proceedings of 2nd International Conference on Mobile Technology, application and systems, 2005, p.1-8.*

[12]. Kozat, U.C.; Kondylis, G.; Ryu, B.; Marina, M. Virtual dynamic backbone for mobile ad hoc networks. *IEEE International Conference on Communication, 2002, p.250-255.*

[13]. Chen, B.; Jamieson, K.; Balakrishnan, H.; Morris, R. SPAN: An energy-efficient coordination algorithm for topology maintenance in ad hoc wireless networks. *In Wireless Networks, 2002, 8,p. 481–494.*

[14]. Krunz, M.; Siam, M. Z.; Nguyen, D. N. Clustering and power management for virtual MIMO communications in wireless sensor networks. *Journal of Ad-hoc network, 2013, 11(5), p. 1571–1587.*

[15]. Naeimi, S.; Ghafghazi, H.; Chow, C.; Ishii, H. Survey on the Taxonomy for Cluster-based Routing Protocols for Homogeneous Wireless Sensor Networks. Sensors. *Journal of Sensor, MDPI, 2012, 12(6),p. 7350–7409.*

[16]. Liu X. A Survey on Clustering Routing Protocols in Wireless Sensor Networks. *Journal of Sensors, 2012, 12,p. 11113-11153*.

[17]. Li, J.; Mohapatra, P. Analytical modeling and mitigation techniques for the energy hole problem in sensor networks. *Pervasive Mobile Computation, 2007, 3, p.233–254.*

[18]. Tran-Quang, V.; Miyoshi, T. A Transmission Range Adjustment Algorithm to Avoid Energy Holes in Wireless Sensor Networks. *Proceedings of 8th Asia-Pacific Symposium on Information and Telecommunication Technologies, Kuching, Malaysia, 2010, p. 15–18.*

[19]. Ishmanov, F.; Malik, A.S.; Kim, A.S. Energy consumption balancing (ECB) issues and mechanisms in Wireless Sensor Networks (WSNs): A comprehensive overview. *Eurosip Transaction on Telecommunication, 2011,22, p.151–167.*

[20]. Karl, H.; Willig, A. Protocols and architectures for wireless sensor network. *A John Wiley and Sons, Ltd, Publication, 2005.*

[21]. Wang, P.; Dui, R.; Akyildiz, I. Collaborative data compression using clustered source coding for wireless multimedia sensor networks. *Proceeding of IEEE Conference on Computer Communications (INFOCOM), San Diego, CA, USA, 2010, p. 1713–1723.*

[22]. Tripathi, R. K.; Singh, Y.N.; Verma, N.K. Two-tiered wireless sensor networks-base station optimal positioning case study. *IET Wireless Sensor Systems, 2012, 2(4),p. 351–360.*

[23]. Mille, M. J.; Vaidya, N. H. A MAC protocol to reduce sensor network energy consumption using a wakeup radio. *IEEE Transaction on Mobile Computing, 2005, 4(3), p.228–242.*

[24]. Zhu, J.; Papavassiliou, S. On the energy-efficient organization and the lifetime of multi-hop sensor networks," *IEEE Communication Letter. 2003, 7(110, p. 537–539.*

[25]. Medagliani, P.; Martalò, M.; Ferrari, G. Clustered Zigbee networks with data fusion: Characterization and performance analysis. *Journal of Ad Hoc Networks, 2011, 9(7),p. 1083-1103.*

[26]. Elfes, A. Occupancy grids: a stochastic spatial representation the active robot perception. *Autonomous Mobile Robots: Perception, Mapping and Navigation, 1991, p. 60-70.*

[27]. Hossain, A.; Chakrabarti, S.; Biswas, P.K. Impact of Sensing Model on Wireless Sensor Network Coverage. *IET wireless sensor system, 2012, 2(3),p. 272-28.*

[28]. Gu,Y.; Wu, Q.; Rao, N. S. V. Optimizing Cluster Heads for Energy Efficiency in Large-Scale Heterogeneous Wireless Sensor Networks. *International Journal of Distributed Sensor Networks, 2010.*

[29]. Dabirmoghaddam, A.; Ghaderi, M.; Williamson, C. Cluster-based correlated data gathering in wireless sensor network. *Proceeding of IEEE International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems , Miami Beach, USA, 2010, p. 163–171.*

[30]. Tuah,N.; Ismail, M.; Jumari, K. Energy efficient algorithm for heterogeneous wireless sensor network. *IEEE International Conference on Control System, Computing and Engineering, 2011, p. 92-96.*

[31]. Heinzelman, W. R.; Chandrakasan, A.; Balakrishnan, H. an application-specific protocol architecture for wireless microsensor networks. *IEEE Transaction on Wireless Communication, 2002, 1(4),p. 660– 670.*

[32]. Depedri, A.; Zanella, R.; Verdone, R. An energy efficient protocol for wireless sensor networks. *Autonomous Intelligent Networks and System, Menlo Park, 2003.*

[33]. Chen, P.; Dea, B. O.; Callaway, E. Energy Efficient System Design with Optimum Transmission Range for Wireless Ad Hoc Networks. *IEEE International Conference on Communication, 2002, 2,p.945-952.*

[34]. Mhatre, V; Rosenberg, C. Design guidelines for wireless sensor networks: communication, clustering and aggregation. *Journal of Ad Hoc Network Journal, 2004, 2(1),p. 45–63.*

[35]. Smaragdakis, G.; Matta, I.; Bestavros, A. SEP: A stable election protocol for clustered heterogeneous wireless sensor networks. *Proceeding of the International Workshop on SANPA, 2004.*

[36]. Pattem, S.; Krishnamachari, B.; Govindan, R. The Impact of Spatial Correlation on Routing with Compression in Wireless Sensor Networks*. ACM/IEEE IPSN, Berkeley, US, 2004, p.28-35.*

[37]. Kim, H.; Kim, S.W.; Lee, S.B.; Son, B. Estimation of the optimal number of cluster-heads in sensor network. *Proceeding of KES, Melbourne, Australia, 2005, 3,p. 87–94.*

[38]. Comeau, F.; Sivakumar, S.C.; Robertson, W.; Phillips W.J. Energy conserving architectures and algorithms for wireless sensor networks. *39th Hawaii International Conference on System Sciences, 2006, p.236c.*

[39]. Chen, H.; Megerian, S. Cluster Sizing and Head Selection for Efficient Data Aggregation and Routing in Sensor Networks. *Proceeding of IEEE WCNC, 2006, p. 2318–2323.*

[40]. Vlajic, N.; Xia, D.Wireless sensor networks: To cluster or not to cluster. *Proceedings of International Symposium on a World of Wireless, Mobile and Multimedia, 2006, 9, p. – 268.*

[41]. Xia, D.; Vlagie, X. Near optimal node clustering in wireless sensor networks for environment monitoring. *21st international conference and applications, 2007,p. 632-641.*

[42]. Yang, H.; Sikdar, B. Optimal Cluster Head Selection in the LEACH Architecture. *Proceeding of IPCCC, 2007, p. 93-100.*

[43]. Wang, H.; Yu, X.; Kong, D.; Yan, X.; Ma, X. Route Protocol of Wireless Sensor Networks Based on Dynamic Setting Cluster. *Proceedings of the International Conference on Information Acquisition, Jeju, Korea, 2007, p. 112–117.*

[44]. Yu, X.; Xu, L.; Wang, H. A Protocol Design of Water in a Remote Measure Based On Mobile Communication Networks and Wireless Sensor Networks. *Automation in Water Resources and Hydrology, 2006, 1, p. 6-10.*

[45]. Wang, Y.; Ma, X.; Xu, L.The designing model o effectively, wireless sensor network data link layer based on information fusion strategy. *Computer Engineering, 2005, 31 23,p. 6-10.*

[46]. Selvakennedy, S.; Sinnappan, S.; Shang, Y. A biologically-inspired clustering protocol for wireless sensor networks. *Journal of Computer Communications, 2007, 30, 14–15, p. 2786–2801.*

[47]. Kumar, D.; Aseri, T.C.; Patel, R.B. EEHC: Energy efficient heterogeneous clustered scheme for wireless sensor networks*. Journal of Computer Communications, 2009, 32, 4,p.662-667.*

[48]. Chan,T.J.; Chen, M.C.; Huang, Y.F.; Lin, J.Y.; Chen, T.R. Optimal Cluster Number Selection in Ad-hoc Wireless Sensor Networks. *WSEAS Transaction on Communication, 2008, 7, 8, 837-846.*

[49]. Chen, H.; Tse, C.K.; Feng, J. Minimizing effective energy consumption in multi-cluster sensor networks for source extraction. *IEEE Transactions on Wireless Communications, 2009, 8(3),p. 1480—1489.*

[50]. Raghuvanshi, S.; Tiwari, S.; Tripathi, R.; Kishor, N. G K Clustering Approach to Determine Optimal Number of Clusters for Wireless Sensor Networks. *Fifth IEEE Conference on Wireless Communication and Sensor network, 2009, p.1-5.*

[51]. Halgamuge, M. N.; Zukerman, M.; Ramamohanarao, K; Vu, H. L. An estimation of sensor energy consumption. *Progress In Electromagnetics Research B, 2009, p.12259-295.*

[52]. Li, H.; Shunjie, X.; Shurong, L.; Weixia, Z.; Zheng, Z. Novel Method for Optimal Number of Cluster Heads in LEACH. *WASE International Conference on Information Engineering, 2009, p. 302-309.*

[53]. Wang, L. C.; Wang, C. W.; Liu, C. M. Optimal number of Clusters in Dense Wireless Sensor Networks: A Cross-Layer Approach*. IEEE Transactions on Vehicular Technology, 2009, 58(2), p.966-976.*

[54]. Kumar, D.; Aseri, T. S.; Patel, R. B. EECHE: Energy efficient cluster head election protocol for heterogeneous Wireless Sensor Networks. *Proceedings of ACM International Conference on Computing, Communication and Control, Bandra, Mumbai, India, 2009, p. 75-80.*

[55]. Li, H.; Shunjie, X.; Guoqiang, W.; Zhe, J. Uneven Virtual Grid-Based Clustering Routing Protocol for Wireless Sensor Networks. *Proceedings of IEEE International Conference on Information and Automation, Zhuhai/Macau, China, 2009, p.397-402.*

[56]. Wang, Y.; Li, C.; Zuo, Y. A selection model for optimal fuzzy clustering algorithm and number of clusters based on competitive comprehensive fuzzy evaluation. *IEEE Transaction on Fuzzy System, 2009, 17(3),568–577.*

[57]. Förster; Förster, A.; Murphy, A. L. Optimal cluster sizes for wireless sensor networks: an experimental analysis. *Ad-Hoc Networks, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, 2010, 28,p. 49–63.*

[58]. Li, W.; Martins, P.; Shen, L. Determination method of optimal number of clusters for clustered wireless sensor networks. *Journal Wireless Communications and Mobile Computing, 2012, 12(2),p. 158 – 168.*

[59]. Ahmed, A.A.; Shi, H.; Shang, Y. A survey on network protocols for wireless sensor networks. *International Conference on Information Technology: Research and Education, 2003, p. 301-305.*

[60]. Raghuvanshi, A. S.; Tiwari, S.; Tripathi, R.; Kishore, N. Optimal number of clusters in wireless sensor networks: a FCM approach. *International Journal of Sensor Networks, 2012, 12(1),p.16-24.*

[61]. Yang, K.; Wu, Y.; Zhou, H. Research of optimal energy consumption model in wirelss sensor network. *2nd International Conference on Computer Engineering and Technology Chengdu, China, 2010, p. V7-421 - V7-424.*

[62]. Navid, A.; Alireza, V.; Wenyao, X.; Mario, G.; Majid, S. Cluster size optimization in sensor networks with decentralized cluster-based protocols. *Journal of Computer Communications, 2012, 35(2),p. 207–220.*

[63]. Chen, B.; Zhang, Y.; Li, Y.; Hao, X.; Fang, Y. A Clustering Algorithm of Cluster-head Optimization for Wireless Sensor Networks Based on Energy. *Journal of Information & Computational Science, 2011,8(11),p. 2129–2136.*

[64]. Tandon, R. Determination of Optimal Number of Clusters in Wireless Sensor Networks. *International Journal of Computer Networks & Communications (IJCNC), 2012, 4(4),p. 235-249.*

[65]. Tuah, N.; Ismail, M.; Jumari, K. Evaluation of Optimal Cluster Size in Heterogeneous Energy Wireless Sensor Network. *1st IEEE International Symposium on Telecommunication Technologies, 2012, p.124 – 130.*

[66]. Tripathi, R.K.; Singh, Y.N.; Verma, N.K. Clustering algorithms for non-uniformly distributed nodes in WSNs. *Electronics Letters, 2013, 49, 4.*

[67]. Dabirmoghaddam, A.; Ghaderi M.; Williamson, C. On the optimal randomized clustering in distributed sensor networks. *Journal of computer network, 2014, 59, 11,p.17–32.*

[68]. Katiyar, V.; Chand, N.; Soni, S. A Survey on clustering algorithms for Heterogeneous Wireless Sensor Networks. *International Journal of Advanced Networking and Applications, 2011, 2(4),p. 745-754.*

[69]. Sheikhpour, R.; Jabbehdari, S.; Khadem-Zadeh, A. Comparison of Energy Efficient Clustering Protocols in Heterogeneous Wireless Sensor Networks. *International Journal of Advanced Science and Technology, 2011, Vol. 36, p.27-40.*

[70]. Mhatre, V.; Rosenberg, C. Homogeneous Vs Heterogeneous Clustered Networks: A Comparative Study . *Proceedings of IEEE ICC, 2004, p.1-6.*

[71]. Shih, E.; Cho, S.; Lee, F. S. B.; Calhoun, H.; Chandrakasan, A. Design Considerations for Energy-Efficient Radios in Wireless Micro-sensor Networks. *Journal of VLSI Signal Process, 2004, 37(1),p. 77–94.*

[72]. Khan, S.; Ahmad, A. Cluster centre initialization algorithm for Kmeans clustering. *Pattern Recognitation Letter*, 2004, 25(1),p. 1293–1302.

[73]. Redmond, S.; Heneghan, C. A method for initializing the *K*-means clustering algorithm using *kd*-trees. *Pattern Recognit. Lett.*, 2007, 28(1), p.965–973

[74]. Huang, H.; Yao, D.; Shen, J.; Ma, K.; Liu, H. Multi-weight based clustering algorithm for wireless sensor networks. *Journal of Electronics & Information Technology, 2008,p. 1489-1492*.

[75]. Stein, E. M.; Shakarchi, R. Real Analysis: Measure Theory, Integration, and Hilbert Spaces. *NJ: Princeton Univ. Press, 2005*.

[76]. Kumar, V.; Dhok, S.B.; Tripathi,R.; Tiwari, S. "Cluster Size Optimization in Gaussian Distributed Wireless Sensor Networks" International Journal of Engineering and Technology (IJET),2014, 6 (3),p. 1581-1592.

[77]. Kumar, V.; and Tiwari, S. "Routing in IPv6 over low-power wireless personal networks (6LowPAN): A survey" Journal of computer Networks and Communication, 2012.

[78]. Kumar, V.; Raghuvansi, A.S.; and Tiwari, S. "Performance study of beacon- enabled IEEE 802.15.4 standards in WSNs with Clustering" International conference on power control and embedded system,2010, p. 1-5.

[79]. Kumar, V.; and Tiwari, S. "Performance of Routing Protocols for Beacon-Enabled IEEE 802.15.4 WSNs with Different Duty Cycle", International Conference on Devices and Communications, 2011, p.1-5.

[80]. Kumar, V.; Raghuvansi, A.S.; and Tiwari, S. "LEACH and Its Derivatives in WSN: A Survey" International Conference on Communication and computational intelligence, 2010/11, p.81-86.

[81]. Kumar, V.; Jain, S.; and Tiwari, S. "Impact of Node Density and Mobility on Scalable Routing Protocols in Mobile AdHoc Networks"Special Issue of International Journal of Computer Applications (0975 8887) on Communication Security, No.5 Mar.2012

[82]. Kumar, V.; Jain, S.; and Tiwari, S. " Performance of Routing Protocols in MANETs with Node Density and Mobility using Omni and Directional Antenna"Special Issue of International Journal of Computer Applications (0975 -8887) on Wireless Communication and Mobile Networks, No.11., ww.ijcaonline.org, Jan.2012,p.51-55.

[83]. Kumar, V.; Dhok, S.B.; Tripathi,R.; Tiwari, S. "A Review Study of Hierarchical Clustering Algorithms for Wireless Sensor Networks" IJCSI International Journal of Computer Science Issues,2014, 11(3), p.92-101.

TNC Transactions on Networks and Communications
SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
VOLUME 2, ISSUE 5
ISSN: 2054 - 7420

TDOA Wireless Localization Comparison Influence of Network Topology

Hao Li and M. Oussalah
University of Birmingham, School of Electronics, Electrical and Computer Engineering, Edgbaston, Birmingham, B15 2TT, UK
hxl441@bham.ac.uk; m.oussalah@bham.ac.uk

ABSTRACT

The interest to wireless positioning techniques has been increasing in recent decades due to wide spread of location-based services as well as constraints imposed by regulator on cellular operator to achieve an accepted level of cellular accuracy regardless of availability of GPS signals. Nevertheless, failure of some base stations cannot be fully avoided, yielding various cellular topologies, which, in turn would likely influence the accuracy of the positioning. This paper explores four types of cellular topologies: balanced, circular, U-shape and linear, which can be inferred from balanced topology structure. Assuming time difference of arrival technology and, up to some extent, time of arrival technology were employed, least square like methods are contrasted with maximum likelihood, Taylor, Chan and hybrid approaches in a simulation platform.

Keywords: wireless positioning, topology, network, TDOA

1 Introduction

With the substantial increase of location based services, which include E911 [1] emergency services where user is tracked with high accuracy using only operator's cellular infrastructure, mapping and path finding, targeted advertising, location based social networking such as MySpace, Friendster or Facebook, the interest to wireless localization techniques has grown drastically in the last two decades. In addition, many ubiquitous applications, including systems like EasyLiving [2] and the Rhino Project [3], among others [4], would benefit from a practical location sensing system. RADAR [5] was one of the first systems to use radio frequency (RF) signal intensity for location-sensing. Small et al. [6] and Smailagic et al. [7] looked at how signal intensity varies over time and developed a location-sensing system based on these observations. Strictly speaking, several localization techniques have been reported in the literature in order to deal with wireless localization, depending on the available technology, which include time-of-arrival (ToA), angle-of-arrival (AOA), time-difference of arrival (TDOA), and received-signal strength (RSS) [8]. Likely the RSS method, where the signal strength from the base station as received in the mobile station is employed as key, which is the less demanding and cheap technology as it does not require any infrastructure change or additional hardware component, which motivates its use in some of above projects like radar [2, 5]. TDOA is recognized for its efficiency and high precision, but requires synchronization among base stations. Indeed, this requires a very accurate timing reference at the mobile which would need to be synchronized with the clock at the base stations. In commonly

DOI: 10.14738/tnc.25.515
Publication Date: 03rd October, 2014
URL: http://dx.doi.org/10.14738/tnc.25.515

employed CDMA system [9], TDOA can be implemented using the pilot tones from different base stations, where the pilot tone transmitted by each cell is used as a coherent carrier reference for synchronization by every mobile in that cell coverage area, which enables the mobile to differentiate each cell site's pilot tone. Therefore the mobile measures the arrival time differences of at least three pilot tones transmitted by three different cells.

Most of the literature survey, including the survey of Guvenc and Chong [8], investigated the performance of the localization algorithms regardless the sensor infrastructure disposition. Although in GSM and UMTS network, it is acknowledgeable that the antenna positioning problem (APP) is one of the major design issues for any mobile operators. It is universally agreed that several factors influence such design. This includes, the (expected) traffic, type of antennas, allocated frequencies, interference, coverage, infrastructure nearby, among others. Since earlier work of Anderson and McGeehan [9] in antenna positioning problem, several other works have been published as well as several national and transnational research projects have been initiated. The idea of integrating several aspects of the network design problem is carried out by Reininger and Caminada [10], as part of the ARNO Project. In the latter, the authors partially relate APP and frequency allocation problem by ‘‘optimizing location and parametrization of the base stations on one shot”.

The integration of locating and configuring base stations is carried further to UMTS networks by Amaldi et al. [11], where the problem of selecting the location and configuring the base stations so as to minimize installation costs as well as to meet the traffic demand is considered. In [12] a trade-off is sought between minimum overlap and desirable cell shapes while the quality of radio coverage is controlled in the constraints. Zimmermann et al. [13] as part of EU ARNO project developed a multi-criteria model that involves a minimum cost, minimum interference and optimum cell shapes. This reveals that most of work in this area has rather been performed from operational research perspective where a multi-criteria decision making like approach has been pursued. Unfortunately less work has been achieved from wireless positioning accuracy perspective has been achieved, although this would significantly contribute towards the E911, for instance. This motivates the current work where some commonly employed techniques involving TDOA and ToA technology are contrasted and investigated with respect to the geometrical disposition of the antennas. More specifically, approximated least square solutions, Maximum likelihood estimation [8], Chan [14], Taylor [15] and a newly introduced combination of Chan-Taylor [16] are compared when considering several antenna topologies. The latter includes linear, circular, U-shape and balanced shapes. Such topology can straightforwardly be inferred from regular (optimal) cellular disposition when some blocking occurs making some BS disabled. The first section of this paper reviews the (eight) main localization techniques employed in this study. Section 3 highlights the simulation platform and comments the obtained results. Finally some conclusive remarks are reported in Section 4.

2 Review of Main TDOA Localization Techniques

Let us consider a general model for the two dimensional (2-D) estimation of a source, consisting of mobile station with Cartesian coordinates (x, y) using M base stations of known locations (X_i, Y_i), i=1 to M. Then the measured distance between the mobile station and the i^{th} base station can be given as:

$$d_i = \sqrt{(X_i - x)^2 + (Y_i - y)^2} + \varepsilon_i = \hat{d}_i + \varepsilon_i = ct_i \quad (1)$$

With $\varepsilon_i \sim \mathcal{N}(\sigma_i^2, 0)$ is the additive white Gaussian noise with variance σ_i^2. $\hat{d}_i$ (i=1, M) stands for estimated distance from MS to i[th] BS, and t_i is the TOA of the signal at the i[th] BS and c is the speed of light. Consequently, for M measurements, the problem comes down to estimating (x,y) from the following set of equations:

$$\begin{bmatrix} (X_1 - x)^2 + (Y_1 - y)^2 \\ . \\ (X_M - x)^2 + (Y_M - y)^2 \end{bmatrix} = \begin{bmatrix} \hat{d}_1^{\,2} \\ . \\ \hat{d}_M^{\,2} \end{bmatrix} \tag{2}$$

2.1 Least Square and Maximum Likelihood Solutions

Assuming that one base station, say r[th] BS, acts as a reference, subtracting r[th] row in (2) from other rows, yields, after some manipulations and defining $K_i = X_i^2 + Y_i^2$ (i=1, M), to matrix equation:

$$AX = \frac{1}{2}B, \tag{3}$$

where

$$A_{[(M-1)\times 2]} = \begin{bmatrix} X_1 - X_r & Y_1 - Y_r \\ X_2 - X_r & Y_2 - Y_r \\ . & . \\ X_M - X_r & Y_M - Y_r \end{bmatrix};\ X = \begin{bmatrix} x \\ y \end{bmatrix};\ B_{[(M-1)\times 1]} = \begin{bmatrix} d_r^2 - d_1^2 - (K_r - K_1) \\ d_r^2 - d_2^2 - (K_r - K_2) \\ . \\ d_r^2 - d_M^2 - (K_r - K_M) \end{bmatrix} \tag{4}$$

A linear least square solution to (4) yields the following *LLT1* solution:

$$X = \frac{1}{2}(A^T A)^{-1} A^T B \tag{5}$$

Another solution proposed in [17] assumes that each BS acts as a servicing BS, and therefore, concatenates the result yielding M (M-1) equations as described by the new *A*, *B* matrices as:

$$A = \begin{bmatrix} X_2 - X_1 & Y_2 - Y_1 \\ X_3 - X_1 & Y_3 - Y_1 \\ \vdots & \vdots \\ X_M - X_1 & Y_M - Y_1 \\ X_3 - X_2 & Y_3 - Y_2 \\ X_4 - X_2 & Y_4 - Y_2 \\ \vdots & \vdots \\ X_M - X_2 & Y_M - y_2 \\ \vdots & \vdots \\ X_M - X_{M-1} & Y_M - Y_{M-1} \end{bmatrix} \quad B = \begin{bmatrix} d_1^2 - d_2^2 - K_1 + K_2 \\ d_1^2 - d_3^2 - K_1 + K_3 \\ \vdots \\ d_1^2 - d_M^2 - K_1 + K_M \\ d_2^2 - d_3^2 - K_2 + K_3 \\ d_2^2 - d_4^2 - K_2 + K_4 \\ \vdots \\ d_2^2 - d_M^2 - K_2 + K_M \\ \vdots \\ d_{M-1}^2 - d_M^2 - K_{M-1} + K_M \end{bmatrix} \tag{6}$$

Where the application of (5) yields what we will refer here as *LLT2* solution

A third approach to least square solution was proposed in [18] where the average of all measurements is subtracted from each measurement equation in (2), yielding new matrices:

$$A=\begin{bmatrix} X_1-\frac{1}{M}\sum_{j=1}^{M}X_j & Y_1-\frac{1}{M}\sum_{j=1}^{M}Y_j \\ X_2-\frac{1}{M}\sum_{j=1}^{M}X_j & Y_1-\frac{1}{M}\sum_{j=1}^{M}Y_j \\ \vdots & \vdots \\ X_M-\frac{1}{M}\sum_{j=1}^{M}X_j & Y_M-\frac{1}{M}\sum_{j=1}^{M}Y_j \end{bmatrix} \quad B=\begin{bmatrix} \frac{1}{M}\sum_{j=1}^{M}d_j^2-d_1^2-\frac{1}{M}\sum_{j=1}^{M}K_j^2+K_1 \\ \frac{1}{M}\sum_{j=1}^{M}d_j^2-d_2^2-\frac{1}{M}\sum_{j=1}^{M}K_j^2+K_2 \\ \vdots \\ \frac{1}{M}\sum_{j=1}^{M}d_j^2-d_M^2-\frac{1}{M}\sum_{j=1}^{M}K_j^2+K_N \end{bmatrix} \tag{7}$$

Again the application of (5) yields a solution referred to as *LLT3*.

A fourth least square solution is obtained when choosing the r[th] reference BS as the one that induces the smallest distance among all other distances but yields same generic solution as (3). Such solution was suggested in [19] and is referred to here as *LLT4*.

The previous least square based solutions discard the knowledge about the uncertainty pervading the measurements (e.g., ε_i) as modelled by the associated variance-covariance matrix, in order to account for such effect, the maximum likelihood solution *MLS* yields as a counterpart of (5) [20]:

$$X=\frac{1}{2}(A^TC^{-1}A)^{-1}A^TC^{-1}B \tag{8}$$

Where A, B are defined as in (4), while the variance-covariance matrix is given by, assuming without loss of generality $\sigma_1=\sigma_2=\ldots=\sigma_M$:

$$C=4d_r{}^2\sigma^2+2\sigma^4+diag\left\{4\sigma^2d_1^2+2\sigma^4 \quad \ldots \quad 4\sigma^2d_i^2+2\sigma^4 \quad \ldots \quad 4\sigma^2d_M^2+2\sigma^4\right\} \tag{9}$$

2.2 Chan and Taylor methods

In Chan's method [14], one assumes the knowledge of the TDOA with respect to a reference BS, say r, so that the measurements are:

$$d_{i,r}=d_i-d_r=cT_{i,r} \tag{10}$$

Where the $T_{i,r}$ is the difference of time arrival between i[th] and r[th] base stations, and d_i are as in (1). Similarly, one denotes $X_{i,r}=X_i-X_r$, $Y_{i,r}=Y_i-Y_r$). Squaring (10) and substituting in (1) yields after some manipulations to [14]:

$$d_{i,r}^2+2d_{i,r}d_r=K_i-K_r-2X_ix-2Y_iy \quad (i=1,M,\ i\neq r) \tag{11}$$

(11) can be put on the form (3) where

$$A_{l(M-1)x3]}=\begin{bmatrix} X_{1,r} & Y_{1,r} & d_{1,r} \\ . & & \\ X_{M,r} & Y_{M,r} & d_{M,r} \end{bmatrix};\ X=\begin{bmatrix} x \\ y \\ d_r \end{bmatrix};\ B_{l(M-1)x1]}=\begin{bmatrix} K_1-K_r-d_{1,r}^2 \\ . \\ K_M-K_r-d_{M,r}^2 \end{bmatrix} \tag{12}$$

Where the unknown vector X contains redundant component d_r, and the solution is approached when first assuming low impact of such dependency to the solution, which is then computed in a two-step strategy. Namely, a linear weighted least square is applied first yielding:

$$X=\left(A^TQ^{-1}A\right)^{-1}A_a{}^TQ^{-1}B \text{ , with } Q=diag\{\sigma_1,\ldots\sigma_M\}. \tag{13}$$

In the second step, the estimate is refined as

$$X = \left(A^T \Psi^{-1} A\right)^{-1} A^T \Psi^{-1} B \tag{14}$$

With

$$\Psi = c^2 BQB \text{ , with } B = diag\{ d_1^0, d_2^0, ..., d_M^0 \} \tag{15}$$

And d_i^0 stands for noise-free estimate of d_i , which is approximated assuming $\mathrm{cov}([x \;\; y \; d_r]^T) \approx (A \Psi^{-1} A)^{-1}$, see [14] for detail.

On the other Taylor's approach [15] to solve (11) in [x, y] starts with an initial guess (x_0; y_0) of the unknown mobile position (x, y), and computes the deviations of the position location estimation:

$$\delta = \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \left(G_t^T Q^{-1} G_t\right)^{-1} G_t^T Q^{-1} h_t \tag{16}$$

With

$$h_t = \begin{bmatrix} d_{2,1} - (d_1 - d_2) \\ d_{3,1} - (d_1 - d_3) \\ . \\ d_{M,1} - (d_1 - d_M) \end{bmatrix}, \quad G_t = \begin{bmatrix} \frac{X_1 - x}{d_1} - \frac{X_2 - x}{d_2} & \frac{Y_1 - y}{d_1} - \frac{Y_2 - y}{d_2} \\ \frac{X_1 - x}{d_1} - \frac{X_3 - x}{d_3} & \frac{Y_1 - y}{d_1} - \frac{Y_3 - y}{d_3} \\ . & \\ . & \\ \frac{X_1 - x}{d_1} - \frac{X_M - x}{d_M} & \frac{Y_1 - y}{d_1} - \frac{Y_M - y}{d_M} \end{bmatrix} \tag{17}$$

In the next iteration, x_0 and y_0 are set to $x_0 + \Delta x$ x and $y_0 + \Delta y$. The whole process is repeated until Δx and Δy are sufficiently small, resulting in the estimated PL of the source (x; y). The Taylor-series method can provide accurate results; however, it requires a close initial guess (x_0, y_0) to guarantee convergence and can be computationally intensive.

In [15], a combination of Chan-Taylor method has been put forward. The proposal assumed a linear combination of the two methods such that the global variance-covariance is minimized. This yield

$$X = \lambda \begin{bmatrix} x_{Chan} \\ y_{Chan} \end{bmatrix} + (1 - \lambda) \begin{bmatrix} x_{Taylor} \\ y_{Taylor} \end{bmatrix}, \tag{18}$$

With

$$\lambda = \frac{P_{Taylor}(1,1)^2 + P_{Taylor}(2,2)^2}{P_{Taylor}(1,1)^2 + P_{Taylor}(2,2)^2 + P_{Chan}(1,1)^2 + P_{Chan}(2,2)^2} \tag{19}$$

Where P_{Taylor} and P_{Chan} stand for variance-covariance matrices associated to Taylor and Chan methods, respectively.

3 Simulation

Similarly to most studies investigating wireless localization techniques, the performances are often evaluated through a set of Monte Carlo simulations. A generic simulation platform is shown in Figure 1. The simulation assumes a set of base station at fixed locations (7 BS in Figure 1). As in practical implementations, the cells have hexagonal shapes in order to restrict the interference between cells as

no overlapping region exists. By abuse, we shall refer to such situation a balanced topology. Nevertheless in case where a blocking occurs in some cells, this yields different topology. For instance if the middle BS in Fig 1 is failed, this yields a circular topology. Similarly if the two first cells in the second row of cells in Fig 1 failed, the cells form a U-like shape, so this is referred to U-shape topology. In total, we shall consider here four different topologies: Circular, U-shape, linear and the balanced one as in Figure 1.

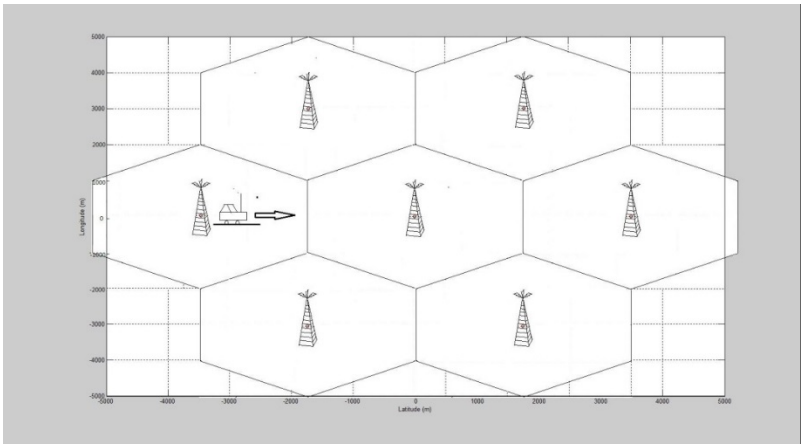


Figure-1: Generic simulation platform (Balanced topology).

Besides we shall consider a vehicle moving at a constant speed in one direction. We therefore, compute for each of the aforementioned localization technique, the localization accuracy with respect to a set of Monte Carlo simulations. The parameters of the simulations for each topology are described in Table 1. The three other topology structures are represented in Figure 2.

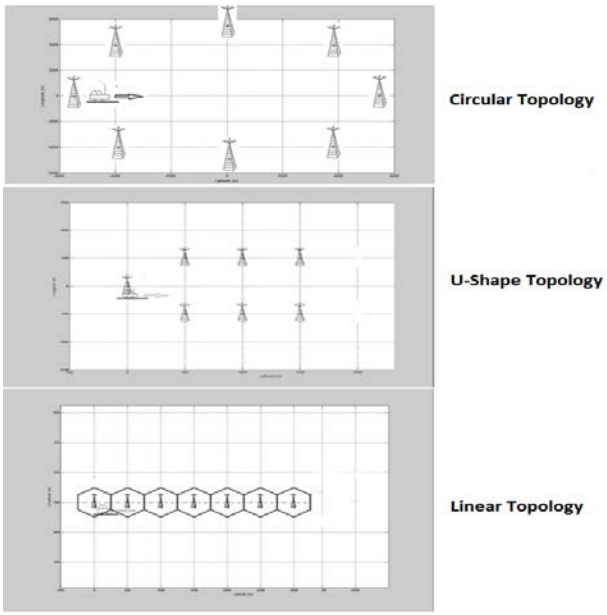


Figure 2: Circular, U- and Linear shape topologies

Typically, to the initial true mobile position is added a random perturbation generated by a zero-mean Gaussian noise with a given standard deviation. A pseudo code highlighting the functioning of the simulation is described in Figure 3.

Table 1: Parameters of the simulation setup

BS Topology	Cell Radius	Noise Standard Deviation	MS Starting Position	Moving Distance	Time	Constant Velocity	Freq. of
Balanced	3000 m	0.1 us	[-5000, 0]	10000 m	50 s	200 m/s	Once / second
Circle	3000 m	0.1 us	[-5000, 0]	10000 m	50 s	200 m/s	Once / second
U-Shape	3000 m	0.1 us	[0, 0]	1500 m	50 s	30 m/s	Once / second
Line	3000 m	0.1 us	[0, 450]	3000 m	50 s	60 m/s	Once / second

```
[MS, RMSE] =LOCATION_ESTIMATION (TOPOLOGY)
RETRIEVE BSi, Vehicle Movement direction, Std σ, Initial MS0
FOR EACH sampling interval k
   FOR EACH Monte Carlo iteration
      MS = ComputePosition (MS0, k)
      Generate a realization of Noise = (0,σ)
      FOR EACH BS
        Calculate distance  d_i = sqrt((BS_i x - MSx)^2 + (BS_i y - MSy)^2) + Noise
       END FOR
      Estimate Position MS= LocationAlgorithm (d, BS, Noise)
    END FOR
   Calculate RMSE of current MS
  END
END
```

Figure 3: Pseudo-code of simulation

In order to quantify the performance of the eight localization techniques, at each sampling interval along the trajectory of the vehicle, the RMSE of the averaged MS estimation over the 1000 Monte Carlo simulations is calculated for each location technique; namely,

$RMSE(t)=\sqrt{\frac{\sum_{i=1}^{n}\left((x_{True}(t)-x_i(t))^2+(y_{True}(t)-y_i(t))^2\right)}{n}}$, where $(x_i(t), y_i(t))$ stands for MS (x, y) estimation at i^{th} Monte Carlo simulation and t sampling interval, and n=1000.

Figures 4, 5, 6 and 7 summarize the localization errors in terms of RMSE of the eight localization techniques when using balanced, circular, U-shape and linear topology.

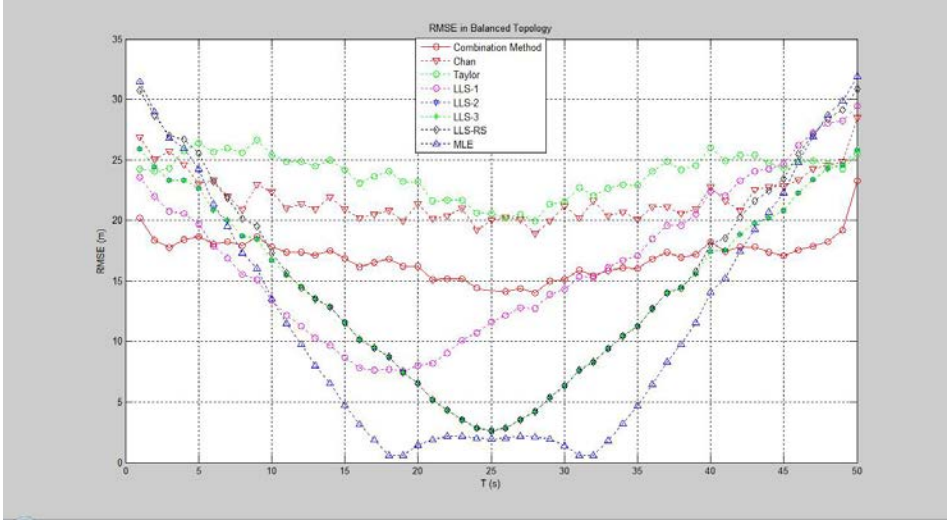


Figure 4: RMSE value in case of balanced topology

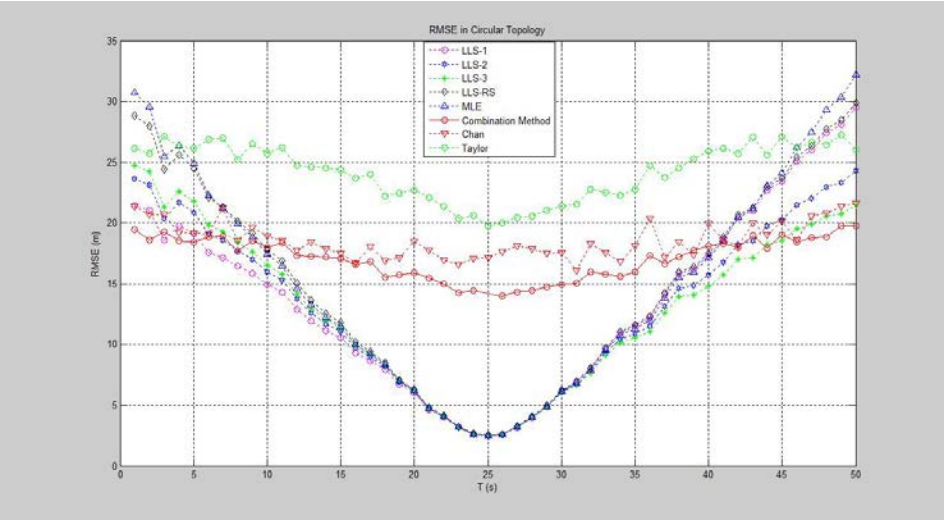


Figure 5: RMSE value in case of Circular topology

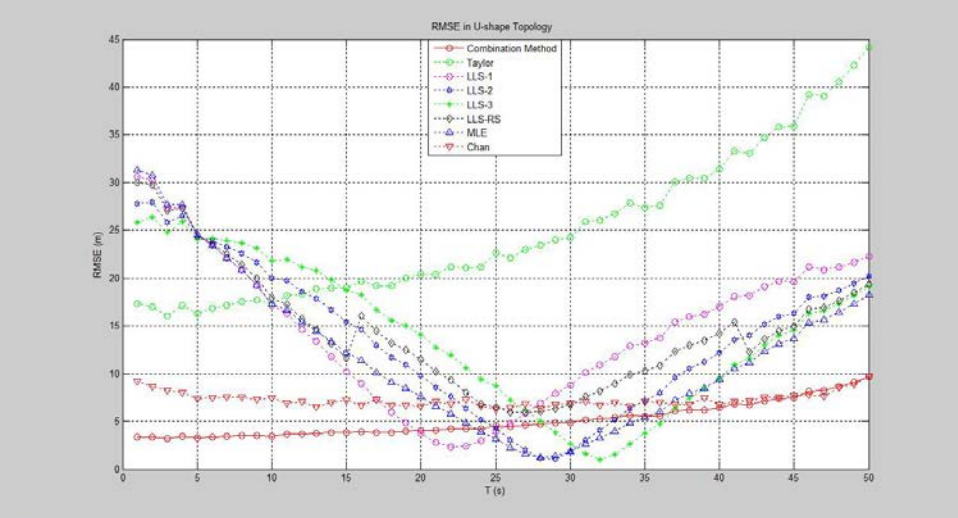


Figure 6: RMSE value in case of U-shape topology

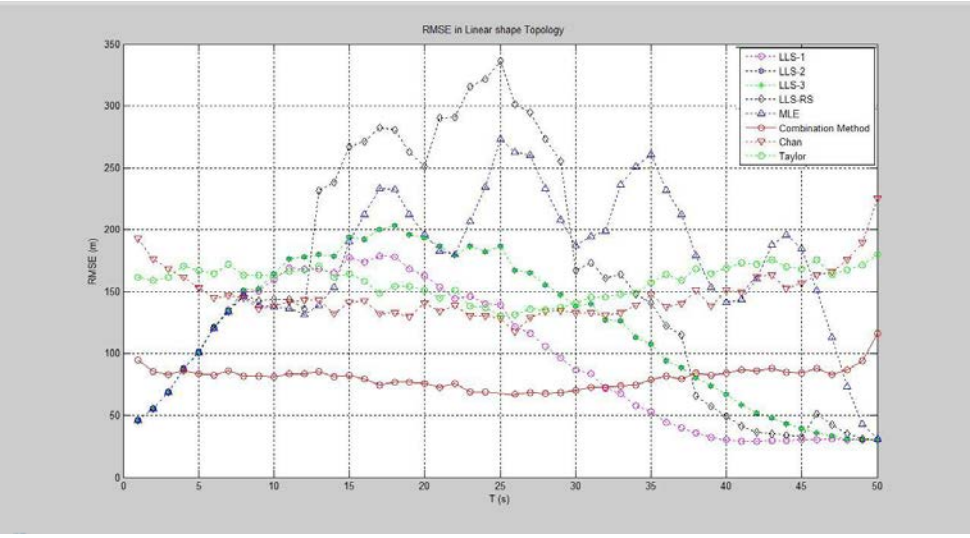


Figure 7: RMSE values in case of Linear shape topology

From the above figures, one can notices the following

- The discrepancy of the various positioning techniques when a change of a topology occurs demonstrates the influence of the topology on the accuracy of the underlying positioning method.
- In the above simulation, at a given sampling interval, the measurements from all base stations are assumed available and aggregated in the localization technique. Although such data cannot be straightforwardly be available in cellular network in practice, where the mobile station is only connected to the base station providing the strongest signal, it is still available from network provider perspective. Besides, such approach is commonly employed in previous work that investigated the performance of cellular/wireless network positioning techniques as testified in the extensive review paper [8].
- Looking at the range of the RMSE values with respect to various topologies reveals that the balanced topology produces the best performance with respect to all positioning techniques, while the linear shape topology yields the worst performance as its associated values RMSE go beyond 340 m as compared to less than 30 m in case of balanced topology. This shows that whenever possible the use of balanced topology should be persuaded. This is mainly due to quality of the obtained measurements, where, at least from geometrical perspective, yields comprehensive intersection of the underlying circles.
- The combination method of Chan and Taylor shows on average that it marginally outperforms the remaining seven topologies regardless the topology employed.
- The investigation of the low values of RMSEs in the above figures reveals that (almost) the least square like methods approach the minimum RMSE value at a sampling time corresponding to the time the vehicle comes close to underlying base station. While such phenomenon is less apparent in case of Chan, Taylor and Combined Chan-Taylor methods where less sensitivity is observed. This is mainly due to the global nature of the above positioning methods.
- The above results have been obtained assuming low noise perturbation as testified by the low standard deviation shown in Table 1. Nevertheless, the influence of the noise intensity cannot

be excluded. On the other hand, few extra simulations with various noise intensities have shown that the generic trends issued from this analysis are not void when the level noise increases. To see it, a 3D graph is depicted in Figure 8 and Figure 9 for balanced and linear like topologies.

- So far, the metric employed for comparison is only related to the accuracy of the positioning technique. Nevertheless, one should bear in mind that some techniques are computationally significantly more expensive than others. From this perspective, LLS1 is computationally the most effective one, and also provides good balance between accuracy and computational cost. While Taylor and combined Chan-Taylor are the most expensive ones because of the iterative approach they do involve. Strictly speaking, even for the LLS1, the computational cost increases with the number of measurements available (value of parameter M). This is mainly due to the cost involved by the matrix inversion operation.

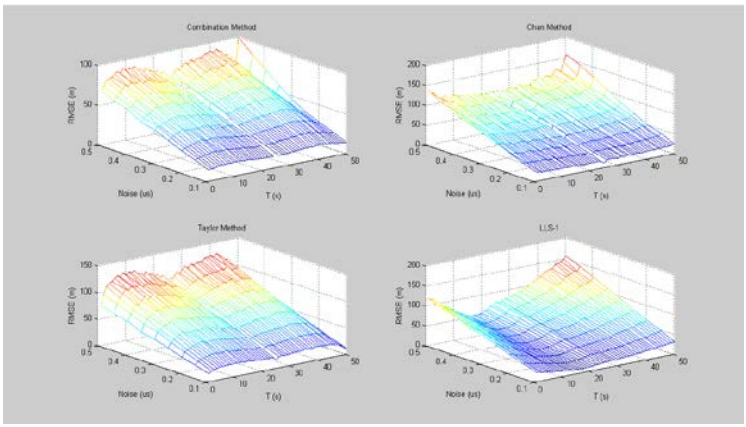


Figure 8: Noise influence in case of balanced topology structure

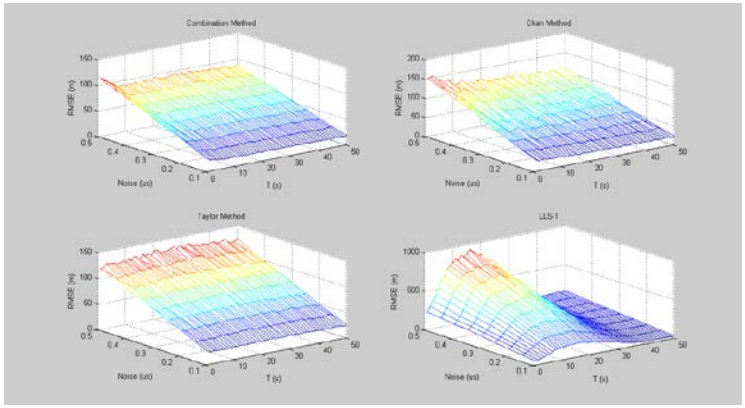


Figure 9: Noise influence in case of Linear shape topology

4 Conclusion

This paper highlights the importance of the antenna positioning when looking at the accuracy of the wireless positioning techniques. Four type of topologies, which can straightforwardly be generated by a regular balanced cellular topology when some blocking occurs, have been investigated. Wireless positioning techniques related to TDOA technology have been examined. This corresponds to four distinct least square based approaches, maximum likelihood, Chan, Taylor and a combined Chan-Taylor method. Simulation results have been obtained assuming a vehicle moving at a constant speed along the given topology. The results demonstrate the credibility of the topology influence on the positioning

accuracy. Besides, the combined Chan-Taylor shows a marginally increased performance in terms of RMSE and sensitivity to base station positioning.

REFERENCES

[1]. C.C. Docket, Revision of the Commission rules to ensure compatibility with enhanced 911 emergency calling systems, RM-8143, Report No. 94-102, FCC, 1994

[2]. B. Brumitt, B. Meyers, J. Krumm, A. Kern, and S. Shafer. EasyLiving: Technologies for intelligent environments, Proceeding of the second international symposium on Handheld and Ubiquitous Computing, Bristol, UK

[3]. G. Abowd, K. Lyons, and K. Scott. The Rhino project, Aug. 1998. http://www.cc.gatech.edu/fce/uvid/rhino.html

[4]. G. D. Abowd, C. G. Atkeson, J. Hong, S. Long, R. Kooper, and M. Pinkerton. Cyberguide: a mobile context-aware tour guide. Wireless Networks, 3(5):421–433, Oct. 1997.

[5]. P. Bahl and V. N. Padmanabhan. Enhancements to the RADAR user location and tracking system. Technical Report MSR-TR-2000-12, Microsoft Research, Feb. 2000

[6]. J. Small, A. Smailagic, and D. P. Siewiorek. Determining user location for context aware computing through the use of a wireless LAN infrastructure, Dec. 2000. http://www-2.cs.cmu.edu/~aura/docdir/small00.pdf

[7]. A. Smailagic, D. Siewiorek, J. Anhalt, D. Kogan, and Y. Wang. Location sensing and privacy in a context aware computing environment. Pervasive Computing, 2001

[8]. I. Guvenc and CC Chong, A survey on TOA Based Wireless Localization and NLOS Mitigation Techniques, IEEE Communication Surveys and Tutorials, 11 (3), 2009, 107-124.

[9]. J. P. McGeehan, and H.R. Anderson. Optimizing Microcell Base Station Locations Using Simulated Annealing Techniques. In: Proc. of the IEEE Vehicular Technology Conference,1994, pp. 858–862

[10]. P. Reininger, and A. Caminada, .Model for GSM Radio Network Optimisation. In: 2nd Intl. ACM/IEEE MobicomWorkshop on Discrete Algorithms and Methods for Mobile Computing and Communications (DIALM), Dallas, December 16, 1998

[11]. E. Amaldi, P. Belotti, A. Capone, F. Malucelli, Optimizing base station location and configuration in UMTS networks. Annals of Operations Research 146, 2006, 135–151.

[12]. A. Jedidi, A., Caminada, A., Finke, G., 2-Objective optimization of cells overlap and geometry with evolutionary algorithms. LectureNotes in Computer Science 3005, 2004, 130–139.

[13]. J. Zimmermann, R. Hons, H. Muhlenbein, ENCON: an evolutionary algorithm for the antenna placement problem. Computers & Industrial Engineering 44, 2003, 209–226

[14]. Y. T., Chan and K.C. Ho, A simple and Efficient Estimator for Hyperbolic Location, IEEE Transactions on Aerospace and Electronic Systems,42(8), 1994, p. 1905-1915.

[15]. R. Shimura and I. Sasase, TDOA mobile terminal positioning with weight control based on received power of pilot symbol in Taylor series estimation, 17th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2006, p. 1 - 5

[16]. Hao Li and M. Oussalah, Combination of Taylor and Chan method in Mobile positioning, in: Proc. of IEEE CIS 2011 Conference, London, pp…

[17]. S. Venkatesh and R. M. Buehrer, "A linear programming approach to NLOS error mitigation in sensor networks," in Proc. IEEE Int. Symp. Information Processing in Sensor Networks (IPSN), Nashville, Tennessee, Apr. 2006, pp. 301–308

[18]. Z. Li, W. Trappe, Y. Zhang, and B. Nath, "Robust statistical methods for securing wireless localization in sensor networks," in Proc. IEEE Int. Symp. Information Processing in Sensor Networks (IPSN), Los Angeles, CA, Apr. 2005, pp. 91–98.

[19]. I. Guvenc, S. Gezici, F. Watanabe, and H. Inamura, "Enhancements to linear least squares localization through reference selection and ML estimation," in Proc. IEEE Wireless Commun. Networking Conf. (WCNC), Las Vegas, NV, Apr. 2008, pp. 284–289

[20]. S. M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory. Upper Saddle River, NJ: Prentice Hall, Inc., 1993

TNC Transactions on Networks and Communications
SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
VOLUME 2, ISSUE 5
ISSN: 2054 - 7420

Cylindrical RF Network Antennas for Coupled Plasma Sources Copper Legs Delayed in Time System Stability Analysis

Ofer Aluf
Physical Electronics Dept., Tel-Aviv University, Ramat-Aviv, Israel.
oferaluf@bezeqint.net

ABSTRACT

In this article, Very Crucial subject discussed cylindrical (closed) RF network antennas for coupled plasma sources copper legs delayed in time system stability analysis. Resonant RF network antennas are important to plasma sources with many applications. The cylindrical resonant RF network antennas run as large volume plasma sources and have stability switching due to system's copper legs parasitic effects. The cylindrical RF network antenna structure is 16-leg cylindrical (Birdcage) RF antenna which has electrical circuit and opposite points of RF feeding and grounding. The vacuum vessel is a glass cylinder closed at the top and bottom by grounding metal plates. Generally there are two popular different resonant RF network assemblies: a cylindrical and a planar RF antenna. The cylindrical RF antenna is built as a high-pass Birdcage coil. The antenna is mounted outside a glass tube. The RF antenna consists of 16 copper legs equally spaced interconnected with capacitors. Due to RF antenna copper leg parasitic effect we get copper leg's current and current derivative with delay $\tau_{1\text{-}k}$ and $\tau_{2\text{-}k}$ (k is leg number index, k=1,…,16). The uncooled antenna is fed at the midpoint and operated with opposite grounded. Alternatively, it can be fed by another transmitter unit. Due to cylindrical antenna parasitic delayed in time, there is a stability issue by analyzing its operation. We consider for simplicity that all copper leg's current parasitic time delayed are equal ($\tau_{1\text{-}1}= \tau_{1\text{-}2}=\ldots =\tau_{1\text{-}16}$) and current derivative parasitic time delayed are equal ($\tau_{2\text{-}1}= \tau_{2\text{-}2}=\ldots =\tau_{2\text{-}16}$). The cylindrical RF network antennas delayed in time equivalent circuit can represent as a delayed differential equations which depend on variable parameters and delays. The investigation of our cylindrical network antenna with copper leg system, a differential equation is based on bifurcation theory [1], a study of possible changes in the structure of the orbits of a delayed differential equation depending on variable parameters. Cylindrical RF network antenna analysis is done under two series of different time delays respect to antenna's copper legs current and current derivative. All of that for optimization of a cylindrical RF network antenna circuit parameter analysis to get the best performance. The cylindrical network antenna with copper leg system can be represented as delayed differential equations which, depending on variable parameters and delays. There is a practical guideline that combines graphical information with analytical work to effectively study the local stability of models involving delay dependent parameters. The stability of a given steady state is determined by the graphs of some function of $\tau_{i\text{-}1}$, ..,τ_{i^-16} (i=1, 2) [2] [3] [4].

Index Terms – Cylindrical RF network antenna, Delay Differential Equations (DDE), Stability, Bifurcation, Orbit.

DOI: 10.14738/tnc.25.570
Publication Date: 16rd October, 2014
URL: http://dx.doi.org/10.14738/tnc.25.570

1 Introduction

In this article, Very Critical and useful subject is discussed: cylindrical (closed) RF network antennas for coupled plasma sources copper legs delayed in time**.** The resonant RF networks can be arranged to form large-area or large-volume plasma sources with properties similar to Inductive Coupled Plasma (ICP) devices. There are medical applications of Birdcage coils and closed and open configurations of the antenna for plasma production are possible and can be analyzed by using mathematical formulation. There are systems of an open network antenna as a large-area planar plasma source and of a closed network antenna as a cylindrical plasma source. Both are composed of similar electrical meshes. Operation at different normal modes shows the capability of this antenna type of large-volume plasma applications.

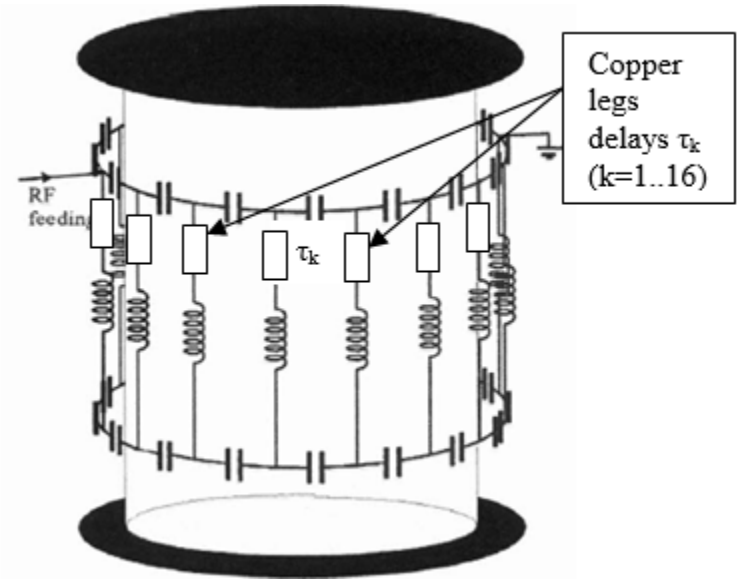


Figure 1. Schematic of the 16-leg cylindrical (Birdcage) RF Network antenna (closed).

An important issue of proper antenna operation is the location of the RF feeding and grounding connections on the antenna. There are a large number of different RF antenna arrangements possible in view of the geometry and RF operation and of plasma obtained. In this paper, we investigated only cylindrical RF antenna which built following a high-pass Birdcage coil. The antenna is mounted outside a glass tube. The RF antenna consists of 16 copper legs (Fig. 1), equally spaced interconnected with capacitors, each copper leg current has parasitic time delayed ($\tau_{1\text{-}1}$… $\tau_{1\text{-}16}$). We consider for simplicity that all copper legs parasitic time delayed are equal ($\tau_{1\text{-}1}$= $\tau_{1\text{-}2}$=… =$\tau_{1\text{-}16}$) and the voltages on delay units (V_ε) are neglected $V_\varepsilon \to \varepsilon$. There is a delay in each Copper leg current $I_1(t-\tau_{1-1}),...,I_{16}(t-\tau_{1-16})$. We consider all interconnected capacitor values are the same (C) and all antenna elements inductance values are the same (L).

$$C_{A1} = C_{A2} = ... = C_{A16} = C \; ; C_{B1} = C_{B2} = ... = C_{B16} = C \tag{1}$$

$$L_1 = L_2 = ... = L_{16} = L \; ; I_{L1} = I_1 \; , I_{L2} = I_2 \; ,..., I_{L16} = I_{16} \tag{2}$$

We choose first case: antenna network is fed by the transmitter unit (S1=OFF, no direct RF feeding). The upper view of 16-leg cylindrical RF antenna network described in Fig. 2.

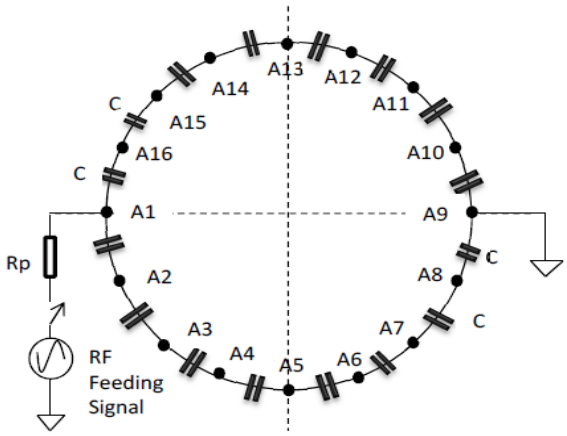


Figure 2. Upper view of 16-leg cylindrical RF antenna.

The lower view of 16-leg cylindrical RF antenna network described in Fig. 3.

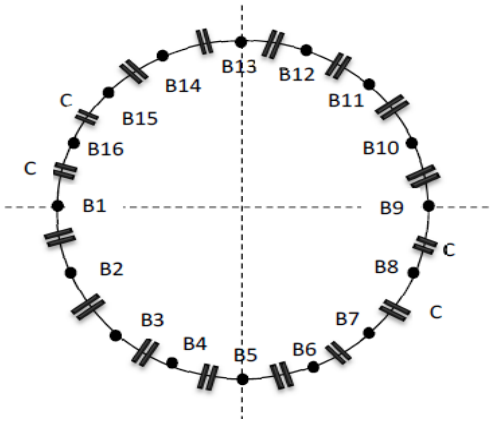


Figure 3. Lower view of 16-leg cylindrical RF antenna

2 Cylindrical RF Network Antennas Equivalent Circuit and represent Delay Differential Equations

Cylindrical RF network antenna system can represent as round strip of capacitors and inductors (Fig. 4a & 4b). The schematic contains RF feeding signal, S1 switch (S1=ON for direct RF signal feeding, S1=OFF for RF signal transmitter feeding). The upper network connecting nodes are A1, A2,…,A16 and the lower network connecting nodes are B1, B2,…,B16. Antenna copper leg current parasitic delays are represented by delay units Tau1-1…Tau1-16 ($\tau_{1-1},..,\tau_{1-16}$). Rp is the parasitic resistance of RF feeding point (A1). The upper system spaced capacitors are CA1,…,CA16 and the lower system spaced capacitors are CB1,…,CB16 [1] [2].

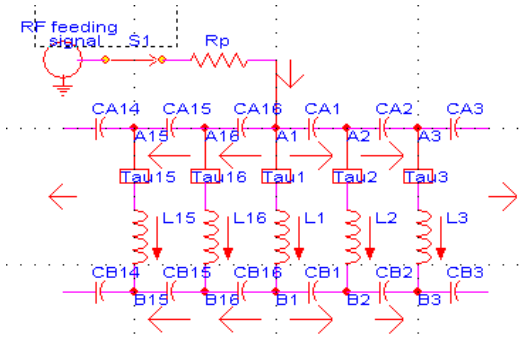


Figure 4a. 16-leg cylindrical RF antenna strip (feeding side)

$I_{CA1} = C_{A1} \cdot \frac{d}{dt}(V_{A1} - V_{A2}), I_{CA2} = C_{A2} \cdot \frac{d}{dt}(V_{A2} - V_{A3})$	3	$I_{CA3} = C_{A3} \cdot \frac{d}{dt}(V_{A3} - V_{A4}), \ldots, I_{CA7} = C_{A7} \cdot \frac{d}{dt}(V_{A7} - V_{A8})$	4
$I_{CA8} = C_{A8} \cdot \frac{dV_{A8}}{dt}$; $I_{CAk} = C_{Ak} \cdot \frac{d}{dt}(V_{Ak} - V_{A(k+1)})$; $k = 1,\ldots,7$	5	$I_{CA16} = C_{A16} \cdot \frac{d}{dt}(V_{A1} - V_{A16}), I_{CA15} = C_{A15} \cdot \frac{d}{dt}(V_{A16} - V_{A15})$ $I_{CA14} = C_{A14} \cdot \frac{d}{dt}(V_{A15} - V_{A14}), \ldots, I_{CA10} = C_{A10} \cdot \frac{d}{dt}(V_{A11} - V_{A10})$	6
$I_{CA9} = C_{A9} \cdot \frac{dV_{A10}}{dt}$; $I_{CAl} = C_{Al} \cdot \frac{d}{dt}(V_{A(l+1)} - V_{Al})$; $l = 10,\ldots,15$	7		

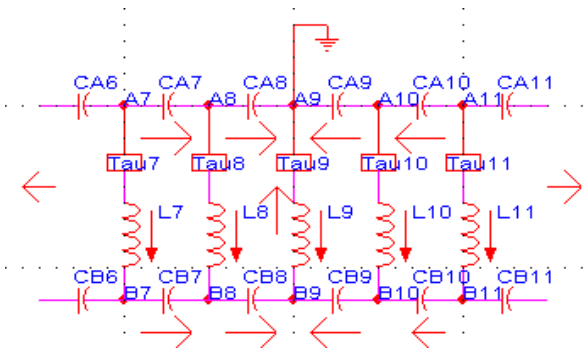


Figure 4b. 16-leg cylindrical RF antenna strip (ground side)

$I_{CB1} = C_{B1} \cdot \frac{d}{dt}(V_{B1} - V_{B2}), I_{CB2} = C_{B2} \cdot \frac{d}{dt}(V_{B2} - V_{B3})$; $k = 1,\ldots,8$ $,\ldots, I_{CB8} = C_{B8} \cdot \frac{d}{dt}(V_{B2} - V_{B3})$; $I_{CBk} = C_{Bk} \cdot \frac{d}{dt}(V_{Bk} - V_{B(k+1)})$	8	$I_{CB16} = C_{B16} \cdot \frac{d}{dt}(V_{B1} - V_{B16}), I_{CB15} = C_{B15} \cdot \frac{d}{dt}(V_{B16} - V_{B15})$ $,\ldots, I_{CB9} = C_{B9} \cdot \frac{d}{dt}(V_{B10} - V_{B9})$; $I_{CBl} = C_{Bl} \cdot \frac{d}{dt}(V_{B(l+1)} - V_{Bl})$ $l = 15,\ldots,9$	9
$V_{A1} - V_{B1} = L_1 \cdot \frac{dI_{L1}}{dt}$; $V_{A2} - V_{B2} = L_2 \cdot \frac{dI_{L2}}{dt}$; $V_{A3} - V_{B3} = L_3 \cdot \frac{dI_{L3}}{dt}$ $,\ldots, V_{A8} - V_{B8} = L_8 \cdot \frac{dI_{L8}}{dt}$; $V_{B9} = L_9 \cdot \frac{dI_{L9}}{dt}$; $V_{A10} - V_{B10} = L_{10} \cdot \frac{dI_{L10}}{dt}$ $,\ldots, V_{A16} - V_{B16} = L_{16} \cdot \frac{dI_{L16}}{dt}$; $V_{Am} - V_{Bm} = L_m \cdot \frac{dI_{Lm}}{dt}$; $m = 1,..,16$; $m \neq 9$	10	$V_{B9} = L_9 \cdot \frac{dI_{L9}}{dt}$; $V_{A9} = 0$; $A9 - ground$	11
$I_{R_p} = I_{CA16} + I_{CA1} + I_{L1}$; $I_{CA1} = I_{CA2} + I_{L2}$; $I_{CA2} = I_{CA3} + I_{L3}$ $,\ldots, I_{CA7} = I_{CA8} + I_{L8}$; $I_{CAl} = I_{CA(l+1)} + I_{L(l+1)}$; $l = 1,\ldots,7$	12	$I_{CA16} = I_{CA15} + I_{L16}$; $I_{CA15} = I_{CA14} + I_{L15}$; $I_{CA14} = I_{CA13} + I_{L14}$ $,\ldots, I_{CA10} = I_{CA9} + I_{L10}$; $I_{CAk} = I_{CA(k-1)} + I_{Lk}$; $k = 16,\ldots,10$	13
$I_{L1} = I_{CB1} + I_{CB16}$; $I_{CB2} = I_{CB1} + I_{L2}$; $I_{CB3} = I_{CB2} + I_{L3}$	14	$I_{CB4} = I_{CB3} + I_{L4}, \ldots, I_{CB8} = I_{CB7} + I_{L8}$; $I_{L9} = I_{CB8} + I_{CB9}$	15
$I_{CBm} = I_{CB(m-1)} + I_{Lm}$; $m = 2,\ldots,8$	16	$I_{CB15} = I_{CB16} + I_{L16}$; $I_{CB14} = I_{CB15} + I_{L15}$; $I_{CB13} = I_{CB14} + I_{L14}$	17
$I_{CB12} = I_{CB13} + I_{L13}, \ldots, I_{CB9} = I_{CB10} + I_{L10}$	18	$I_{CBn} = I_{CB(n+1)} + I_{L(n+1)}$; $n = 15,\ldots,9$	19

Upon mathematic manipulation we get the following expressions:

$\frac{I_{C_{A1}}}{C_{A1}}-\frac{I_{C_{B1}}}{C_{B1}}=L_1\cdot\frac{d^2I_{L1}}{dt^2}-L_2\cdot\frac{d^2I_{L2}}{dt^2}\ ;\ C_{A1}=C_{B1}=C$	20	$L_1=L_2=L\ ;\ \frac{1}{LC}\cdot(I_{C_{A1}}-I_{C_{B1}})=\frac{d^2I_{L1}}{dt^2}-\frac{d^2I_{L2}}{dt^2}$	21
$\frac{I_{C_{A2}}}{C_{A2}}-\frac{I_{C_{B2}}}{C_{B2}}=L_2\cdot\frac{d^2I_{L2}}{dt^2}-L_3\cdot\frac{d^2I_{L3}}{dt^2}\ ;\ C_{A2}=C_{B2}=C$	22	$L_2=L_3=L\ ;\ \frac{1}{LC}\cdot(I_{C_{A2}}-I_{C_{B2}})=\frac{d^2I_{L2}}{dt^2}-\frac{d^2I_{L3}}{dt^2}$	23
$\cdots\cdots\frac{I_{C_{A7}}}{C_{A7}}-\frac{I_{C_{B7}}}{C_{B7}}=L_7\cdot\frac{d^2I_{L7}}{dt^2}-L_8\cdot\frac{d^2I_{L8}}{dt^2}$	24	$L_7=L_8=L\ ;\ \frac{1}{LC}\cdot(I_{C_{A7}}-I_{C_{B7}})=\frac{d^2I_{L7}}{dt^2}-\frac{d^2I_{L8}}{dt^2}$	25
$C_{A7}=C_{B7}=C\ ;L_1=L_2=\ldots=L_{16}=L\ ;k=1,\ldots,7$	26	$\frac{1}{LC}\cdot(I_{C_{Ak}}-I_{C_{Bk}})=\frac{d^2I_{Lk}}{dt^2}-\frac{d^2I_{L(k+1)}}{dt^2}\ ;\ k=1,\ldots,7$	27
$\frac{1}{LC}\cdot(I_{C_{A10}}-I_{C_{B10}})=\frac{d^2I_{L11}}{dt^2}-\frac{d^2I_{L10}}{dt^2}\ ;\ C_{A10}=C_{B10}=C$	28	$\cdots\frac{1}{LC}\cdot(I_{C_{A15}}-I_{C_{B15}})=\frac{d^2I_{L16}}{dt^2}-\frac{d^2I_{L15}}{dt^2}\ ;\ C_{A15}=C_{B15}=C$	29
$\frac{1}{LC}\cdot(I_{C_{Am}}-I_{C_{Bm}})=\frac{d^2I_{L(m+1)}}{dt^2}-\frac{d^2I_{Lm}}{dt^2}\ ;\ m=10,\ldots,15$	30	$C_{A8}=C_{B8}=C\ ;V_{A9}=0\ ;\ \frac{1}{LC}\cdot(I_{C_{A8}}-I_{C_{B8}})=\frac{d^2I_{L8}}{dt^2}+\frac{d^2I_{L9}}{dt^2}$	31
$C_{A9}=C_{B9}=C\ ;\ V_{A9}=0\ ;\ \frac{1}{LC}\cdot(I_{C_{A9}}-I_{C_{B9}})=\frac{d^2I_{L9}}{dt^2}+\frac{d^2I_{L10}}{dt^2}$	32	$C_{A16}=C_{B16}=C\ ;\ V_{A9}=0\ ;\ \frac{1}{LC}\cdot(I_{C_{A16}}-I_{C_{B16}})=\frac{d^2I_{L1}}{dt^2}-\frac{d^2I_{L16}}{dt^2}$	33
$I_{R_P}=I_{C_{A16}}+I_{C_{A1}}+I_{L1}\ ;\ I_{C_{A1}}=I_{C_{A2}}+I_{L2}\ ;\ I_{C_{A2}}=I_{C_{A3}}+I_{L3}$	34	$I_{C_{A3}}=I_{C_{A4}}+I_{L4}\ ;\ I_{C_{A4}}=I_{C_{A5}}+I_{L5}\ ;\ I_{C_{A5}}=I_{C_{A6}}+I_{L6}$	35
$I_{C_{A6}}=I_{C_{A7}}+I_{L7}\ ;\ I_{C_{A7}}=I_{C_{A8}}+I_{L8}\ ;\ I_{C_{A10}}=I_{C_{A9}}+I_{L10}$	36	$I_{C_{A11}}=I_{C_{A10}}+I_{L11}\ ;\ I_{C_{A12}}=I_{C_{A11}}+I_{L12}\ ;\ I_{C_{A13}}=I_{C_{A12}}+I_{L13}$ $I_{C_{A14}}=I_{C_{A13}}+I_{L14}\ ;\ I_{C_{A15}}=I_{C_{A14}}+I_{L15}\ ;\ I_{C_{A16}}=I_{C_{A15}}+I_{L16}$ $I_{L1}=I_{C_{B1}}+I_{C_{B16}}\ ;\ I_{L9}=I_{C_{B8}}+I_{C_{B9}}\ ;\ I_{C_{B2}}=I_{C_{B1}}+I_{L2}$	37
$I_{C_{B3}}=I_{C_{B2}}+I_{L3}\ ;\ I_{C_{B4}}=I_{C_{B3}}+I_{L4}\ ;\ I_{C_{B5}}=I_{C_{B4}}+I_{L5}$	38	$I_{C_{B6}}=I_{C_{B5}}+I_{L6}\ ;\ I_{C_{B7}}=I_{C_{B6}}+I_{L7}\ ;\ I_{C_{B8}}=I_{C_{B7}}+I_{L8}$	39
$I_{C_{B9}}=I_{C_{B10}}+I_{L10}\ ;\ I_{C_{B10}}=I_{C_{B11}}+I_{L11}\ ;\ I_{C_{B11}}=I_{C_{B12}}+I_{L12}$	40	$I_{C_{B12}}=I_{C_{B13}}+I_{L13}\ ;\ I_{C_{B13}}=I_{C_{B14}}+I_{L14}\ ;\ I_{C_{B14}}=I_{C_{B15}}+I_{L15}$ $I_{C_{B15}}=I_{C_{B16}}+I_{L16}$	41

S1 is OFF for RF signal transmitter feeding.

$I_{R_P}=0\Rightarrow I_{C_{A16}}+I_{C_{A1}}+I_{L1}=0$	42	$I_{C_{A1}}=I_{C_{A8}}+\sum_{k=2}^{8}I_{Lk}\ ;\ I_{C_{A2}}=I_{C_{A8}}+\sum_{k=3}^{8}I_{Lk}\ ;\ I_{C_{A3}}=I_{C_{A8}}+\sum_{k=4}^{8}I_{Lk}$	43
$I_{C_{A4}}=I_{C_{A8}}+\sum_{k=5}^{8}I_{Lk}\ ;\ I_{C_{A5}}=I_{C_{A8}}+\sum_{k=6}^{8}I_{Lk}\ ;\ I_{C_{A6}}=I_{C_{A8}}+\sum_{k=7}^{8}I_{Lk}$	44	$I_{C_{A7}}=I_{C_{A8}}+I_{L8}\ ;\ I_{C_{A16}}=I_{C_{A9}}+\sum_{k=10}^{16}I_{Lk}\ ;\ I_{C_{A15}}=I_{C_{A9}}+\sum_{k=10}^{15}I_{Lk}$	45
$I_{C_{A14}}=I_{C_{A9}}+\sum_{k=10}^{14}I_{Lk}\ ;\ I_{C_{A13}}=I_{C_{A9}}+\sum_{k=10}^{13}I_{Lk}\ ;\ I_{C_{A12}}=I_{C_{A9}}+\sum_{k=10}^{12}I_{Lk}$	46	$I_{C_{A11}}=I_{C_{A9}}+\sum_{k=10}^{11}I_{Lk}\ ;\ I_{C_{A10}}=I_{C_{A9}}+I_{L_{10}}$	47
$I_{C_{B1}}=I_{L9}-I_{C_{B16}}-\sum_{k=2,k\neq9}^{16}I_{Lk}\ ;\ I_{C_{B2}}=I_{L9}-I_{C_{B16}}-\sum_{k=3,k\neq9}^{16}I_{Lk}$	48	$I_{C_{B3}}=I_{L9}-I_{C_{B16}}-\sum_{k=4,k\neq9}^{16}I_{Lk}\ ;\ I_{C_{B4}}=I_{L9}-I_{C_{B16}}-\sum_{k=5,k\neq9}^{16}I_{Lk}$	49
$I_{C_{B5}}=I_{L9}-I_{C_{B16}}-\sum_{k=6,k\neq9}^{16}I_{Lk}\ ;\ I_{C_{B6}}=I_{L9}-I_{C_{B16}}-\sum_{k=7,k\neq9}^{16}I_{Lk}$	50	$I_{C_{B7}}=I_{L9}-I_{C_{B16}}-\sum_{k=8,k\neq9}^{16}I_{Lk}\ ;\ I_{C_{B8}}=I_{L9}-I_{C_{B16}}-\sum_{k=10}^{16}I_{Lk}$	51
$I_{C_{B9}}=I_{C_{B16}}+\sum_{k=10}^{16}I_{Lk}\ ;\ I_{C_{B10}}=I_{C_{B16}}+\sum_{k=11}^{16}I_{Lk}$	52	$I_{C_{B11}}=I_{C_{B16}}+\sum_{k=12}^{16}I_{Lk}\ ;\ I_{C_{B12}}=I_{C_{B16}}+\sum_{k=13}^{16}I_{Lk}$	53
$I_{C_{B13}}=I_{C_{B16}}+\sum_{k=14}^{16}I_{Lk}\ ;\ I_{C_{B14}}=I_{C_{B16}}+\sum_{k=15}^{16}I_{Lk}$	54	$I_{C_{B15}}=I_{C_{B16}}+I_{L16}\ ;\ I_{L1}=I_{C_{B1}}+I_{C_{B16}}$	55

We get the following additional exppressions:

$\frac{2}{L\cdot C}\cdot(I_{L2}-I_{L4})=\frac{d^2I_{L1}}{dt^2}-\frac{d^2I_{L5}}{dt^2}-2\cdot[\frac{d^2I_{L2}}{dt^2}-\frac{d^2I_{L4}}{dt^2}]$	56	$\frac{2}{L\cdot C}\cdot(I_{L6}-I_{L8})=\frac{d^2I_{L5}}{dt^2}+\frac{d^2I_{L9}}{dt^2}-2\cdot[\frac{d^2I_{L6}}{dt^2}-\frac{d^2I_{L8}}{dt^2}]$	57
$\frac{-2}{L\cdot C}\cdot(I_{L10}-I_{L12})=\frac{d^2I_{L9}}{dt^2}+\frac{d^2I_{L13}}{dt^2}+2\cdot[\frac{d^2I_{L10}}{dt^2}-\frac{d^2I_{L12}}{dt^2}]$	58	$\frac{-2}{L\cdot C}\cdot(I_{L14}-I_{L16})=\frac{d^2I_{L1}}{dt^2}-\frac{d^2I_{L13}}{dt^2}+2\cdot[\frac{d^2I_{L14}}{dt^2}-\frac{d^2I_{L16}}{dt^2}]$	60

We add the first and second above equations:

$$[*]\quad \frac{2}{L\cdot C}\cdot\{(I_{L2}-I_{L4})+(I_{L6}-I_{L8})\}=\frac{d^2I_{L1}}{dt^2}+\frac{d^2I_{L9}}{dt^2}-2\cdot[\frac{d^2I_{L2}}{dt^2}-\frac{d^2I_{L4}}{dt^2}+\frac{d^2I_{L6}}{dt^2}-\frac{d^2I_{L8}}{dt^2}] \tag{60}$$

We add the third and forth above equations:

$$[**]\quad \frac{-2}{L\cdot C}\cdot\{(I_{L10}-I_{L12})+(I_{L14}-I_{L16})\}=\frac{d^2I_{L1}}{dt^2}+\frac{d^2I_{L9}}{dt^2}+2\cdot[\frac{d^2I_{L10}}{dt^2}-\frac{d^2I_{L12}}{dt^2}+\frac{d^2I_{L14}}{dt^2}-\frac{d^2I_{L16}}{dt^2}] \tag{61}$$

Integrating the last two results ([**]-[*]) gives the following:

$$\frac{-2}{L\cdot C}\cdot\{I_{L10}-I_{L12}+I_{L14}-I_{L16}+I_{L2}-I_{L4}+I_{L6}-I_{L8}\}=2\cdot[\frac{d^2I_{L10}}{dt^2}-\frac{d^2I_{L12}}{dt^2}+\frac{d^2I_{L14}}{dt^2}-\frac{d^2I_{L16}}{dt^2}+\frac{d^2I_{L2}}{dt^2}-\frac{d^2I_{L4}}{dt^2}+\frac{d^2I_{L6}}{dt^2}-\frac{d^2I_{L8}}{dt^2}] \tag{62}$$

We define new global variables for our Cylindrical RF network antennas system.

$$Y=I_{L10}-I_{L12}+I_{L14}-I_{L16}+I_{L2}-I_{L4}+I_{L6}-I_{L8} \tag{63}$$

$$X=\frac{dI_{L10}}{dt}-\frac{dI_{L12}}{dt}+\frac{dI_{L14}}{dt}-\frac{dI_{L16}}{dt}+\frac{dI_{L2}}{dt}-\frac{dI_{L4}}{dt}+\frac{dI_{L6}}{dt}-\frac{dI_{L8}}{dt}\ ;\ \frac{dY}{dt}=X\ ;\ \frac{dX}{dt}=\frac{-1}{L\cdot C}\cdot Y \tag{64}$$

Due to RF antenna copper leg parasitic effect, we get copper leg's current and current derivative with delay $\tau_{1\text{-}k}$ and $\tau_{2\text{-}k}$ (k is leg number index, k=1,…,16). We consider for simplicity

$\tau_{1\text{-}1}=\tau_{1\text{-}2}=\ldots=\tau_{1\text{-}16}$; $\tau_{2\text{-}1}=\tau_{2\text{-}2}=\ldots=\tau_{2\text{-}16}$. $I_{Lk}(t)\to I_{Lk}(t-\tau_{1\text{-}k})$ $I'_{Lk}(t)=\frac{dI_{Lk}(t)}{dt}$; $I'_{Lk}(t)\to I'_{Lk}(t-\tau_{2-k})$.

We consider no delay effect on

$$\frac{dI'_{Lk}(t)}{dt}.\ Y(t)\to Y(t-\tau_1);X(t)\to X(t-\tau_2). \tag{65}$$

$\tau_1=\tau_{1\text{-}1}=\tau_{1\text{-}2}=\ldots=\tau_{1\text{-}16}$; $\tau_2=\tau_{2\text{-}1}=\tau_{2\text{-}2}=\ldots=\tau_{2\text{-}16}$.

$$\frac{dY}{dt}=X(t-\tau_2)\ ;\ \frac{dX}{dt}=\frac{-1}{L\cdot C}\cdot Y(t-\tau_1) \tag{66}$$

To find the Equilibrium points (fixed points) of the Cylindrical RF network antennas system is by

$$\lim_{t\to\infty}Y(t-\tau_1)=Y(t)\ \text{and}\ \lim_{t\to\infty}X(t-\tau_2)=X(t)\ .\ \frac{dY}{dt}=0\ ;\ \frac{dX}{dt}=0 \tag{67}$$

$$\forall\ t\gg\tau_1\ ;t\gg\tau_2\ \ \exists\ (t-\tau_1)\approx t\ ;(t-\tau_2)\approx t\ ,t\to\infty \tag{68}$$

We get two equations and the only fixed point is

$$E^{(0)}(Y^{(0)},X^{(0)})=(0,0)\,. \tag{69}$$

$$Y^{(0)}=I^{(0)}_{L10}-I^{(0)}_{L12}+I^{(0)}_{L14}-I^{(0)}_{L16}+I^{(0)}_{L2}-I^{(0)}_{L4}+I^{(0)}_{L6}-I^{(0)}_{L8}=0 \tag{70}$$

$$X^{(0)}=I'^{(0)}_{L10}-I'^{(0)}_{L12}+I'^{(0)}_{L14}-I'^{(0)}_{L16}+I'^{(0)}_{L2}-I'^{(0)}_{L4}+I'^{(0)}_{L6}-I'^{(0)}_{L8}=0 \tag{71}$$

Stability analysis : The standard local stability analysis about any one of the equilibrium points of Cylindrical RF network antennas system consists in adding to coordinates [Y X] arbitrarily small increments of exponential form $[y\ x]\cdot e^{\lambda\cdot t}$, and retaining the first order terms in y, x. The system of two homogeneous equations leads to a polynomial characteristics equation in the Eigen values λ . The polynomial characteristics equations accept by set the below current and current derivative respect to time into two Cylindrical RF network antennas system equations. Cylindrical RF network antennas system fixed values with arbitrarily small increments of exponential form [y x].$e^{\lambda.t}$ are: i=0 (first fixed point), i=1 (second fixed point), i=2 (third fixed point).

$$\begin{aligned}&Y(t)=Y^{(i)}+y\cdot e^{\lambda\cdot t}\ ;X=X^{(i)}+x\cdot e^{\lambda\cdot t}\ ;Y(t-\tau_1)=Y^{(i)}+y\cdot e^{\lambda\cdot(t-\tau_1)}\\&X(t-\tau_2)=X^{(i)}+x\cdot e^{\lambda\cdot(t-\tau_2)}\ \ \forall\ i=0,\ 1,\ 2\end{aligned} \tag{72}$$

We choose the above expressions for our $Y(t),\ X(t)$ as small displacement [y x] from the system fixed points at time t=0.

$$Y(t=0)=Y^{(i)}+y\ ;X(t=0)=X^{(i)}+x \tag{73}$$

for $\lambda<0,\ t>0$ the selected fixed point is stable otherwise $\lambda>0,\ t>0$ is Unstable. Our Cylindrical RF network antennas system tend to the selected fixed point exponentially for $\lambda<0,\ t>0$ otherwise go away from the selected fixed point exponentially. λ is the eigenvalue parameter which establish if the fixed point is stable or unstable, additionally his absolute value ($|\lambda|$) establish the speed of flow toward or away from the selected fixed point [1] [2].

Table-1. Cylindrical RF network antennas system eigenvalues options

	λ<0	**λ>0**
t=0	$Y(t=0)= Y^{(i)}+y$ $X(t=0)= X^{(i)}+x$	$Y(t=0)= Y^{(i)}+y$ $X(t=0)= X^{(i)}+x$
t>0	$Y(t)= Y^{(i)}+y\cdot e^{-\lvert\lambda\rvert\bullet t}$ $X(t)= X^{(i)}+x\cdot e^{-\lvert\lambda\rvert\bullet t}$	$Y(t)= Y^{(i)}+y\cdot e^{\lvert\lambda\rvert\bullet t}$ $X(t)= X^{(i)}+x\cdot e^{\lvert\lambda\rvert\bullet t}$
t→∞	$Y(t\to\infty)= Y^{(i)}$ $X(t\to\infty)= X^{(i)}$	$Y(t\to\infty, \lambda>0)\sim ye^{\lvert\lambda\rvert\bullet t}$ $X(t\to\infty, \lambda>0)\sim xe^{\lvert\lambda\rvert\bullet t}$

The speeds of flow toward or away from the selected fixed point for Cylindrical RF network antennas system currents and currents derivative respect to time are

$$\frac{dY(t)}{dt}=\lim_{\Delta t\to 0}\frac{Y(t+\Delta t)-Y(t)}{\Delta t}=\lim_{\Delta t\to 0}\frac{Y^{(i)}+y\cdot e^{\lambda\cdot(t+\Delta t)}-[Y^{(i)}+y\cdot e^{\lambda\cdot t}]}{\Delta t} \tag{74}$$
$$=\lim_{\Delta t\to 0}\frac{y\cdot e^{\lambda\cdot t}\cdot[e^{\lambda\cdot\Delta t}-1]}{\Delta t}\xrightarrow{e^{\lambda\cdot\Delta t}\approx 1+\lambda\cdot\Delta t}\lim_{\Delta t\to 0}\frac{y\cdot e^{\lambda\cdot t}\cdot[1+\lambda\cdot\Delta t-1]}{\Delta t}=\lambda\cdot y\cdot e^{\lambda\cdot t}$$

$$\frac{dX(t)}{dt}=\lim_{\Delta t\to 0}\frac{X(t+\Delta t)-X(t)}{\Delta t}=\lim_{\Delta t\to 0}\frac{X^{(i)}+x\cdot e^{\lambda\cdot(t+\Delta t)}-[X^{(i)}+x\cdot e^{\lambda\cdot t}]}{\Delta t} \tag{75}$$
$$=\lim_{\Delta t\to 0}\frac{x\cdot e^{\lambda\cdot t}\cdot[e^{\lambda\cdot\Delta t}-1]}{\Delta t}\xrightarrow{e^{\lambda\cdot\Delta t}\approx 1+\lambda\cdot\Delta t}\lim_{\Delta t\to 0}\frac{x\cdot e^{\lambda\cdot t}\cdot[1+\lambda\cdot\Delta t-1]}{\Delta t}=\lambda\cdot x\cdot e^{\lambda\cdot t}$$

and the time derivative of the above equations:

$$\frac{dY(t)}{dt}=y\cdot\lambda\cdot e^{\lambda\cdot t}\ ;\ \frac{dX(t)}{dt}=x\cdot\lambda\cdot e^{\lambda\cdot t} \tag{76}$$

$$\frac{dY(t-\tau_1)}{dt}=y\cdot\lambda\cdot e^{\lambda\cdot(t-\tau_1)}=y\cdot\lambda\cdot e^{\lambda\cdot t}\cdot e^{-\tau_1\cdot\lambda}$$
$$\frac{dX(t-\tau_2)}{dt}=x\cdot\lambda\cdot e^{\lambda\cdot(t-\tau_2)}=x\cdot\lambda\cdot e^{\lambda\cdot t}\cdot e^{-\tau_2\cdot\lambda} \tag{77}$$

First we take the Cylindrical RF network antennas (Y) differential equation: $\frac{dY}{dt}=X$ and adding to it coordinates [Y X] arbitrarily small increments of exponential form $[y\ x]\cdot e^{\lambda\cdot t}$ and retaining the first order terms in y, x .

$$\lambda\cdot y\cdot e^{\lambda\cdot t}=X^{(i)}+x\cdot e^{\lambda\cdot t}\ ;\ X^{(i=0)}=0\ ;\ \lambda_1=\frac{x}{y}\approx 1>0 \tag{78}$$

Second we take the Cylindrical RF network antennas (X) differential equation: $\frac{dX}{dt}=\frac{-1}{L\cdot C}\cdot Y$ and adding to it coordinates [Y X] arbitrarily small increments of exponential form $[y\ x]\cdot e^{\lambda\cdot t}$ and retaining the first order terms in y, x .

$$\lambda\cdot x\cdot e^{\lambda\cdot t}=\frac{-1}{L\cdot C}\cdot[Y^{(i)}+y\cdot e^{\lambda\cdot t}]\ ;\ Y^{(i=0)}=0\ ;\ \lambda_2=\frac{-1}{L\cdot C}\cdot\frac{y}{x}\Big|_{\frac{y}{x}\approx 1}=\frac{-1}{L\cdot C}<0 \tag{79}$$

we have saddle fixed point otherwise it is unstable node (both eigenvalues are positive). We define $Y(t-\tau_1)=Y^{(i)}+y\cdot e^{\lambda\cdot(t-\tau_1)}$ $X(t-\tau_2)=X^{(i)}+x\cdot e^{\lambda\cdot(t-\tau_2)}$ then we get two delayed differential equations

respect to adding to it coordinates [Y X] arbitrarily small increments of exponential form $[y\ x]\cdot e^{\lambda\cdot t}$. In the equilibrium points: $Y^{(0)}=0\ ;\ X^{(0)}=0$

$$\begin{aligned}&\lambda\cdot y\cdot e^{\lambda\cdot t}=X^{(0)}+x\cdot e^{\lambda\cdot(t-\tau_2)}\ ;\ X^{(0)}=0\Rightarrow\lambda\cdot y=x\cdot e^{-\lambda\cdot\tau_2}\\&\lambda\cdot x\cdot e^{\lambda\cdot t}=\frac{-1}{L\cdot C}\cdot[Y^{(0)}+y\cdot e^{\lambda\cdot(t-\tau_1)}]\ ;\ Y^{(0)}=0\Rightarrow\\&\lambda\cdot x=\frac{-1}{L\cdot C}\cdot y\cdot e^{-\lambda\cdot\tau_1}\end{aligned}\tag{80}$$

We get the following set of eigenvalues equations:

$$-\lambda\cdot y+x\cdot e^{-\lambda\cdot\tau_2}=0\ ;\ \frac{-1}{L\cdot C}\cdot y\cdot e^{-\lambda\cdot\tau_1}-\lambda\cdot x=0\tag{81}$$

The small increments Jacobian of our Cylindrical RF network antennas.

$$\begin{pmatrix}-\lambda & e^{-\lambda\cdot\tau_2}\\ \frac{-1}{L\cdot C}\cdot e^{-\lambda\cdot\tau_1} & -\lambda\end{pmatrix}\cdot\begin{pmatrix}y\\x\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}\tag{82}$$

$$A-\lambda\cdot I=\begin{pmatrix}-\lambda & e^{-\lambda\cdot\tau_2}\\ \frac{-1}{L\cdot C}\cdot e^{-\lambda\cdot\tau_1} & -\lambda\end{pmatrix}\ ;\ \det|A-\lambda\cdot I|=0\tag{83}$$

$$D(\lambda,\tau_1,\tau_2)=\lambda^2+\frac{1}{L\cdot C}\cdot e^{-\lambda\cdot\tau_1}\cdot e^{-\lambda\cdot\tau_2}\tag{84}$$

We have three stability analysis cases: $\tau_1=\tau\ ;\ \tau_2=0$ or $\tau_2=\tau\ ;\ \tau_1=0$ or $\tau_1=\tau_2=\tau$ otherwise $\tau_1\neq\tau_2$. We need to get characteristics equations as all above stability analysis cases. We study the occurrence of any possible stability switching resulting from the increase of value of the time delay τ for the general characteristic equation $D(\lambda,\tau)$.

$$D(\lambda,\tau)=P_n(\lambda,\tau)+Q_m(\lambda,\tau)\cdot e^{-\lambda\tau}\tag{85}$$

The expression for $P_n(\lambda,\tau)$ is

$$\begin{aligned}&P_n(\lambda,\tau)=\sum_{k=0}^{n}P_k(\tau)\cdot\lambda^k=P_0(\tau)+P_1(\tau)\cdot\lambda+P_2(\tau)\cdot\lambda^2\\&+P_3(\tau)\cdot\lambda^3+.....\end{aligned}\tag{86}$$

The expression for $Q_m(\lambda,\tau)$ is

$$Q_m(\lambda,\tau)=\sum_{k=0}^{m}q_k(\tau)\cdot\lambda^k=q_0(\tau)+q_1(\tau)\cdot\lambda+q_2(\tau)\cdot\lambda^2+\ldots\ldots \tag{87}$$

3 Cylindrical RF Network Antennas System Second Order Characteristic Equation $\tau_1=\tau\ ;\ \tau_2=0$ & $\tau_1=0\ ;\ \tau_2=\tau$

The first case we analyze is when there is delay in Cylindrical RF network antennas leg's current and no delay in antennas leg's current derivative or opposite [4] [5].

$$D(\lambda,\tau_1=0,\tau_2)=\lambda^2+\frac{1}{L\cdot C}\cdot e^{-\lambda\cdot\tau_2}{}_{\tau_2=\tau}=\lambda^2+\frac{1}{L\cdot C}\cdot e^{-\lambda\cdot\tau} \tag{88}$$

$$D(\lambda,\tau_1,\tau_2=0)=\lambda^2+\frac{1}{L\cdot C}\cdot e^{-\lambda\cdot\tau_1}|_{\tau_1=\tau}=\lambda^2+\frac{1}{L\cdot C}\cdot e^{-\lambda\cdot\tau} \tag{89}$$

$$D(\lambda,\tau)=P_n(\lambda,\tau)+Q_m(\lambda,\tau)\cdot e^{-\lambda\tau} \tag{90}$$

The expression for $P_n(\lambda,\tau)$ is

$$P_n(\lambda,\tau)=\sum_{k=0}^{n}P_k(\tau)\cdot\lambda^k=P_0(\tau)+P_1(\tau)\cdot\lambda+P_2(\tau)\cdot\lambda^2 \tag{91}$$
$$=\lambda^2\ ;\ P_2(\tau)=1\ ;\ P_1(\tau)=0\ ;\ P_0(\tau)=0$$

The expression for $Q_m(\lambda,\tau)$ is

$$Q_m(\lambda,\tau)=\sum_{k=0}^{m}q_k(\tau)\cdot\lambda^k=q_0(\tau)=\frac{1}{L\cdot C} \tag{92}$$

Our Cylindrical RF network antennas system second order characteristic equation:

$$D(\lambda,\tau)=\lambda^2+a(\tau)\cdot\lambda+b(\tau)\cdot\lambda\cdot e^{-\lambda\cdot\tau}+c(\tau)+d(\tau)\cdot e^{-\lambda\cdot\tau} \tag{93}$$

$$\text{Then } a(\tau)=0\ ;\ \mathrm{b}(\tau)=0\ ;\ \mathrm{c}(\tau)=0\ ;\ \mathrm{d}(\tau)=\frac{1}{L\cdot C} \tag{94}$$

$\tau\in R_{+0}$ and $a(\tau),\mathrm{b}(\tau),\mathrm{c}(\tau),\mathrm{d}(\tau):R_{+0}\to R$ are differentiable functions of class $C^1(R_{+0})$ such that $\mathrm{c}(\tau)+\mathrm{d}(\tau)=\frac{1}{L\cdot C}\neq 0$ for all $\tau\in R_{+0}$ and for any $\tau,\mathrm{b}(\tau),\mathrm{d}(\tau)$ are not simultaneously zero. We have

$$P(\lambda,\tau)=P_n(\lambda,\tau)=\lambda^2+a(\tau)\cdot\lambda+c(\lambda)=\lambda^2 \tag{95}$$

$$Q(\lambda,\tau)=Q_m(\lambda,\tau)=b(\tau)\cdot\lambda+d(\tau)=\frac{1}{L\cdot C} \tag{96}$$

We assume that $P_n(\lambda,\tau)=P_n(\lambda)$ and $Q_m(\lambda,\tau)=Q_m(\lambda)$ can't have common imaginary roots. That is for any real number

$$\omega \; ; \; p_n(\lambda = i\cdot\omega,\tau) + Q_m(\lambda = i\cdot\omega,\tau) \neq 0 \tag{97}$$

$$-\omega^2 + \frac{1}{L\cdot C} \neq 0 \tag{98}$$

$$F(\omega,\tau) = |P(i\cdot\omega,\tau)|^2 - |Q(i\cdot\omega,\tau)|^2 \quad F(\omega,\tau) = \omega^4 - \frac{1}{(L\cdot C)^2}$$
$$= (c-\omega^2)^2 + \omega^2\cdot a^2 - (\omega^2\cdot b^2 + d^2) \tag{99}$$

Hence

$$F(\omega,\tau) = 0 \text{ implies } \omega^4 - \frac{1}{(L\cdot C)^2} = 0 \tag{100}$$

And its roots are given by

$$\omega_+^2 = \frac{1}{2}\cdot\{(b^2 + 2\cdot c - a^2) + \sqrt{\Delta}\} = \frac{\sqrt{\Delta}}{2} \tag{101}$$

$$\omega_-^2 = \frac{1}{2}\cdot\{(b^2 + 2\cdot c - a^2) - \sqrt{\Delta}\} = -\frac{\sqrt{\Delta}}{2} \tag{102}$$

$$\Delta = (b^2 + 2\cdot c - a^2) - 4\cdot(c^2 - d^2) = \frac{4}{L^2\cdot C^2} \tag{103}$$

Therefore the following holds:

$$2\cdot\omega_{+/-}^2 - (b^2 + 2\cdot c - a^2) = \pm\sqrt{\Delta} \; ; \; 2\cdot\omega_{+/-}^2 = \pm\sqrt{\Delta} \tag{104}$$

$$\text{Furthermore } P_R(i\cdot\omega,\tau) = c(\tau) - \omega^2(\tau) = -\omega^2(\tau) \tag{105}$$

$$P_I(i\cdot\omega,\tau) = \omega(\tau)\cdot a(\tau) = 0 \; ; \; Q_R(i\cdot\omega,\tau) = d(\tau) = \frac{1}{L\cdot C} \tag{106}$$

$$Q_I(i\cdot\omega,\tau) = \omega(\tau)\cdot b(\tau) = 0 \text{ hence} \tag{107}$$

$$\sin\theta(\tau) = \frac{-P_R(i\cdot\omega,\tau)\cdot Q_I(i\cdot\omega,\tau) + P_I(i\cdot\omega,\tau)\cdot Q_R(i\cdot\omega,\tau)}{|Q(i\cdot\omega,\tau)|^2} \tag{108}$$

$$\cos\theta(\tau) = -\frac{P_R(i\cdot\omega,\tau)\cdot Q_R(i\cdot\omega,\tau) + P_I(i\cdot\omega,\tau)\cdot Q_I(i\cdot\omega,\tau)}{|Q(i\cdot\omega,\tau)|^2} \tag{109}$$

$$\sin\theta(\tau) = \frac{-(c-\omega^2)\cdot\omega\cdot b + \omega\cdot a\cdot d}{\omega^2\cdot b^2 + d^2} = 0 \tag{110}$$

$$\cos\theta(\tau) = -\frac{(c-\omega^2)\cdot d + \omega^2 \cdot a \cdot b}{\omega^2 \cdot b^2 + d^2} = \omega^2 \cdot L \cdot C \tag{111}$$

Which jointly with $\omega^4 - \frac{1}{(L\cdot C)^2} = 0$ (112)

Defines the maps $S_n(\tau) = \tau - \tau_n(\tau)$; $\tau \in \mathrm{I}$, $\mathrm{n} \in \mathbb{N}_0$ (113)

That are continuous and differentiable in τ based on Lema 1.1 (see Appendix A). Hence we use theorem 1.2 (see Appendix B). This prove the theorem 1.3 (see Appendix C) and theorem 1.4 (see Appendix D).

Remark: a, b, c, d parameters are independent of delay parameter τ even we use $a(\tau), \mathrm{b}(\tau), \mathrm{c}(\tau), \mathrm{d}(\tau)$.

4 CYLINDRICAL RF NETWORK ANTENNAS SYSTEM SECOND ORDER CHARACTERISTIC EQUATION $\tau_1 = \tau$; $\tau_2 = \tau$

The second case we analyze is when there is delay both in Cylindrical RF network antennas leg's current and current time derivative [4] [5].

$$D(\lambda, \tau_1 = \tau, \tau_2 = \tau) = \lambda^2 + \frac{1}{L\cdot C}\cdot e^{-\lambda\cdot\tau}\cdot e^{-\lambda\cdot\tau} \tag{114}$$

$$D(\lambda,\tau) = P_n(\lambda,\tau) + Q_m(\lambda,\tau)\cdot e^{-\lambda\tau} \tag{115}$$

The expression for $P_n(\lambda,\tau)$ is

$$P_n(\lambda,\tau) = \sum_{k=0}^{n} P_k(\tau)\cdot\lambda^k = P_0(\tau) + P_1(\tau)\cdot\lambda + P_2(\tau)\cdot\lambda^2 = \lambda^2 \tag{116}$$
$$P_2(\tau) = 1 \; ; \; P_1(\tau) = 0 \; ; \; P_0(\tau) = 0$$

The expression for

$$Q_m(\lambda,\tau) \; ; \; Q_m(\lambda,\tau) = \sum_{k=0}^{m} q_k(\tau)\cdot\lambda^k = \frac{1}{L\cdot C}\cdot e^{-\lambda\cdot\tau} \tag{117}$$

Taylor expansion : $e^{-\lambda\tau} \approx 1 - \lambda\cdot\tau + \frac{\lambda^2\cdot\tau^2}{2}$ since we need

n > m [BK] analysis we choose $e^{-\lambda\tau} \approx 1 - \lambda\cdot\tau$ then we get $Q_m(\lambda,\tau) = \sum_{k=0}^{m} q_k(\tau)\cdot\lambda^k = \frac{1}{L\cdot C}\cdot(1-\lambda\cdot\tau) = \frac{1}{L\cdot C} - \frac{1}{L\cdot C}\cdot\lambda\cdot\tau$ (118)

$$q_0(\tau,\lambda) = \frac{1}{L\cdot C} \; ; \; q_1(\tau) = -\frac{1}{L\cdot C}\cdot\tau \; ; \; q_2(\tau) = 0 \tag{119}$$

Our Cylindrical RF network antennas system second order characteristic equation:

$$D(\lambda,\tau) = \lambda^2 + a(\tau)\cdot\lambda + b(\tau)\cdot\lambda\cdot e^{-\lambda\cdot\tau} + c(\tau) + d(\tau)\cdot e^{-\lambda\cdot\tau} \tag{120}$$

Then

$$a(\tau)=0\ ;\ b(\tau)=\frac{-1}{L\cdot C}\cdot\tau\ ;\ c(\tau)=0\ ;\ d(\tau)=\frac{1}{L\cdot C} \tag{121}$$

And in the same manner like our previous case analysis:

$$P(\lambda,\tau)=P_n(\lambda,\tau)=\lambda^2 \tag{122}$$

$$Q(\lambda,\tau)=Q_m(\lambda,\tau)=\frac{1}{L\cdot C}-\frac{1}{L\cdot C}\cdot\lambda\cdot\tau \tag{123}$$

we assume that $P_n(\lambda,\tau)=P_n(\lambda)$ and $Q_m(\lambda,\tau)$ can't have common imaginary roots. That is for any real number

$$\omega\ ;\ p_n(\lambda=i\cdot\omega,\tau)+Q_m(\lambda=i\cdot\omega,\tau)\neq 0 \tag{124}$$

$$-\omega^2-i\cdot\omega\cdot\frac{1}{L\cdot C}\cdot\tau+\frac{1}{L\cdot C}\neq 0 \tag{125}$$

$$F(\omega,\tau)=\mid P(i\cdot\omega,\tau)\mid^2-\mid Q(i\cdot\omega,\tau)\mid^2\ ;\ P(i\cdot\omega,\tau)=-\omega^2 \tag{126}$$

$$P_R(i\cdot\omega,\tau)=-\omega^2\ ;\ P_I(i\cdot\omega,\tau)=0 \tag{127}$$

$$Q(\lambda=i\cdot\omega,\tau)=-i\cdot\omega\cdot\frac{1}{L\cdot C}\cdot\tau+\frac{1}{L\cdot C} \tag{128}$$

$$Q_I(\lambda=i\cdot\omega,\tau)=-\omega\cdot\frac{1}{L\cdot C}\cdot\tau\ ;\ Q_R(\lambda=i\cdot\omega,\tau)=\frac{1}{L\cdot C} \tag{129}$$

$$\mid P(i\cdot\omega,\tau)\mid^2=P_I^2+P_R^2\ \ ;\ \ \mid Q(i\cdot\omega,\tau)\mid^2=Q_I^2+Q_R^2 \tag{130}$$

$$\mid P(i\cdot\omega,\tau)\mid^2=P_I^2+P_R^2=\omega^4 \tag{131}$$

$$\mid Q(i\cdot\omega,\tau)\mid^2=\omega^2\cdot\frac{\tau^2}{(L\cdot C)^2}+\frac{1}{(L\cdot C)^2} \tag{132}$$

$$F(\omega,\tau)=\omega^4-\omega^2\cdot\frac{\tau^2}{(L\cdot C)^2}-\frac{1}{(L\cdot C)^2} \tag{133}$$

Hence

$$F(\omega,\tau)=0 \text{ implies } \omega^4-\omega^2\cdot\frac{\tau^2}{(L\cdot C)^2}-\frac{1}{(L\cdot C)^2}=0 \tag{134}$$

$$F_\omega=4\cdot\omega^3-2\cdot\omega\cdot\frac{\tau^2}{(L\cdot C)^2}=2\cdot\omega\cdot[2\cdot\omega^2-\frac{\tau^2}{(L\cdot C)^2}] \tag{135}$$

$$F_\tau=\frac{-\omega^2\cdot 2\cdot\tau}{(L\cdot C)^2} \tag{136}$$

$$P_{I\omega}=0\ ;\ P_{R\omega}=-2\cdot\omega\ ;\ Q_{I\omega}=-\frac{\tau}{L\cdot C}$$
$$Q_{R\omega}=0\ ;\ P_{I\tau}=0\ ;\ P_{R\tau}=0\ ;\ Q_{R\tau}=0\ ;\ Q_{I\tau}=-\frac{\omega}{L\cdot C} \quad (137)$$

The expressions for U, V can be derive easily [BK] : $x=\tau$

$$U=(P_R\cdot P_{I\omega}-P_I\cdot P_{R\omega})-(Q_R\cdot Q_{I\omega}-Q_I\cdot Q_{R\omega}) \quad (138)$$

$$V=(P_R\cdot P_{Ix}-P_I\cdot P_{Rx})-(Q_R\cdot Q_{Ix}-Q_I\cdot Q_{Rx}) \quad (139)$$

$$V=\frac{\omega}{L^2\cdot C^2}\ \ ;\ U=\frac{\tau}{L^2\cdot C^2} \quad (140)$$

$$\omega_\tau=-\frac{F_\tau}{F_\omega}$$

and we get the expression:

$$\omega_\tau=-\frac{\frac{-\omega^2\cdot 2\cdot\tau}{(L\cdot C)^2}}{2\cdot\omega\cdot[2\cdot\omega^2-\frac{\tau^2}{(L\cdot C)^2}]}=\frac{\frac{-\omega\cdot\tau}{(L\cdot C)^2}}{[2\cdot\omega^2-\frac{\tau^2}{(L\cdot C)^2}]} \quad (141)$$

Defines the maps $S_n(\tau)=\tau-\tau_n(\tau)\ ;\ \tau\in \mathrm{I}\ ,\ \mathrm{n}\in\mathbb{N}_0$

Defines the maps $S_n(\tau)=\tau-\tau_n(\tau)\ ;\ \tau\in \mathrm{I}\ ,\ \mathrm{n}\in\mathbb{N}_0$

That are continuous and differentiable in τ based on Lema 1.1 (see Appendix A). Hence we use theorem 1.2 (see Appendix B). This prove the theorem 1.3 (see Appendix C) and theorem 1.4 (see Appendix D).

Remark: Taylor approximation for $e^{-\lambda\tau}\approx 1-\lambda\cdot\tau$ gives us Good stability analysis approximation only for restricted delay time interval.

5 CYLINDRICAL RF NETWORK ANTENNAS SYSTEM STABILITY ANALYSIS UNDER DELAYED VARIABLES IN TIME $\tau_1=\tau\ ;\ \tau_2=\tau$

Our Cylindrical RF network antennas homogeneous system for y, x leads to a characteristic equation for the eigenvalue λ having the form $P(\lambda)+Q(\lambda)\cdot e^{-\lambda\cdot\tau}=0$; second case $\tau_1=\tau\ ;\ \tau_2=\tau$

$$D(\lambda,\tau_1=\tau,\ \tau_2=\tau)=\lambda^2+\frac{1}{L\cdot C}\cdot e^{-\lambda\cdot\tau}\cdot e^{-\lambda\cdot\tau} \quad (142)$$

We estimate $e^{-\lambda\tau}\approx 1-\lambda\cdot\tau$

$$D(\lambda,\tau_1=\tau,\ \tau_2=\tau)=\lambda^2+\frac{1}{L\cdot C}\cdot(1-\lambda\cdot\tau)\cdot e^{-\lambda\cdot\tau} \quad (143)$$

$$D(\lambda,\tau_1=\tau,\ \tau_2=\tau)=\lambda^2+(-\lambda\cdot\frac{1}{L\cdot C}\cdot\tau+\frac{1}{L\cdot C})\cdot e^{-\lambda\cdot\tau} \tag{144}$$

We use different parameters terminology from our last characteristics parameters definition:

$$k\to j\ ;\ p_k(\tau)\to a_j\ ;\ q_k(\tau)\to c_j\ ;\ n=2\ ;\ m=1 \tag{145}$$

Additionally $P_n(\lambda,\tau)\to P(\lambda)$; $Q_m(\lambda,\tau)\to Q(\lambda)$

$$\text{then } P(\lambda)=\sum_{j=0}^{2}a_j\cdot\lambda^j \text{ and } Q(\lambda)=\sum_{j=0}^{1}c_j\cdot\lambda^j . \tag{146}$$

$$P(\lambda)=\lambda^2;\ Q(\lambda,\tau)=-\lambda\cdot\frac{1}{L\cdot C}\cdot\tau+\frac{1}{L\cdot C} \tag{147}$$

$n,m\in\mathbb{N}_0$, $n>m$ and $a_j,c_j:\mathrm{R}_{+0}\to R$ are continuous and differentiable function of τ such that $a_0+c_0\neq 0$. In the following "—"denotes complex and conjugate. $P(\lambda),Q(\lambda)$

Are analytic functions in λ and differentiable in τ .

And the coefficients: $\{a_j(C,L),c_j(C,L,\tau)\}\in\mathbb{R}$ depend on Cylindrical RF network antennas C, L, τ values.

$$a_0=0,\ a_1=0,\ a_2=1\ ;\ c_0=\frac{1}{L\cdot C},\ c_1=-\frac{1}{L\cdot C}\cdot\tau \tag{148}$$

Unless strictly necessary, the designation of the variation arguments (C,L,τ) will subsequently be omitted from P, Q, a_j, c_j. The coefficients a_j, c_j are continuous, and differentiable functions of their arguments, and direct substitution shows that

$$a_0+c_0=\frac{1}{L\cdot C}\neq 0\ ;\ \frac{1}{L\cdot C}\neq 0\ \forall\ C,L,\tau\in\mathbb{R}_+ \text{ i.e} \tag{149}$$

$\lambda=0$ is not a root of characteristic equation. Furthermore $P(\lambda)$, $Q(\lambda)$ are analytic function of λ for which the following requirements of the analysis (see kuang, 1993, section 3.4) can also be verified in the present case [4] [5].

a) If $\lambda=i\cdot\omega$, $\omega\in\mathbb{R}$ then $P(i\cdot\omega)+Q(i\cdot\omega)\neq 0$, i.e P and Q have no common imaginary roots. This condition was verified numerically in the entire (C,L,τ) domain of interest.

b) $|Q(\lambda)/P(\lambda)|$ is bounded for $|\lambda|\to\infty$, $\mathrm{Re}\,\lambda\geq 0$. No roots bifurcation from ∞. Indeed, in the limit

$$|Q(\lambda)/P(\lambda)|=|\frac{-\lambda\cdot\frac{1}{L\cdot C}\cdot\tau+\frac{1}{L\cdot C}}{\lambda^2}| \tag{150}$$

c) $$F(\omega) = |P(i\cdot\omega)|^2 - |Q(i\cdot\omega)|^2 \quad (151)$$

$$F(\omega,\tau) = \omega^4 - \omega^2\cdot\frac{\tau^2}{(L\cdot C)^2} - \frac{1}{(L\cdot C)^2} \quad (152)$$

Has at most a finite number of zeros. Indeed, this is a bi-cubic polynomial in ω (second degree in ω^2).

d) Each positive root $\omega(C,L,\tau)$ of $F(\omega)=0$ is continuous and differentiable with respect to C, L, τ. This condition can only be assessed numerically.

In addition, since the coefficients in P and Q are real, we have $\overline{P(-i\cdot\omega)} = P(i\cdot\omega)$, and $\overline{Q(-i\cdot\omega)} = Q(i\cdot\omega)$ thus $\lambda = i\cdot\omega$, $\omega > 0$ may be on eigenvalue of characteristic equation. The analysis consists in identifying the roots of characteristic equation situated on the imaginary axis of the complex λ – plane, where by increasing the parameters C, L and delay τ, Reλ may, at the crossing ,Change its sign from (-) to (+), i.e. from stable focus $E^{(0)}(Y^{(0)}, X^{(0)}) = (0,0)$ to an unstable one, or vice versa. This feature may be further assessed by examining the sign of the partial derivatives with respect to C, L and antenna parameters.

$$\wedge^{-1}(C) = \left(\frac{\partial \operatorname{Re}\lambda}{\partial C}\right)_{\lambda=i\cdot\omega},\ L,\tau = const \quad (153)$$

$$\wedge^{-1}(L) = \left(\frac{\partial \operatorname{Re}\lambda}{\partial L}\right)_{\lambda=i\cdot\omega},\ C,\tau = const \quad (154)$$

$$\wedge^{-1}(\tau) = \left(\frac{\partial \operatorname{Re}\lambda}{\partial \tau}\right)_{\lambda=i\cdot\omega},\ C, L, \text{where } \omega \in \mathbb{R}_+. \quad (155)$$

For the first case $\tau_1 = \tau$; $\tau_2 = \tau$ we get the following results

$$P_R(i\cdot\omega) = -\omega^2\ ;\ P_I(i\cdot\omega) = 0\ ;\ Q_R(i\cdot\omega) = \frac{1}{L\cdot C} \quad (156)$$

$$Q_I(i\cdot\omega) = \frac{-\omega\cdot\tau}{L\cdot C};\ F(\omega) = 0 \text{ yield to} \quad (157)$$

$$\omega^4 - \omega^2\cdot\frac{\tau^2}{(L\cdot C)^2} - \frac{1}{(L\cdot C)^2} = 0\ ;\ \chi^2 = \omega^4\ ;\ \chi = \omega^2 \quad (158)$$

$$\chi^2 - \chi\cdot\frac{\tau^2}{(L\cdot C)^2} - \frac{1}{(L\cdot C)^2} = 0 \quad (159)$$

$$\chi = \frac{\tau^2}{2\cdot(L\cdot C)^2} \pm \frac{1}{2}\cdot\sqrt{\frac{\tau^4}{(L\cdot C)^4} + 4\cdot\frac{1}{(L\cdot C)^2}} \quad (160)$$

$$\chi = \omega^2 \Rightarrow \omega = \pm\sqrt{\frac{\tau^2}{2\cdot(L\cdot C)^2} \pm \frac{1}{2}\cdot\sqrt{\frac{\tau^4}{(L\cdot C)^4} + 4\cdot\frac{1}{(L\cdot C)^2}}} \tag{161}$$

$$\frac{\tau^4}{(L\cdot C)^4} + 4\cdot\frac{1}{(L\cdot C)^2} > 0$$

always and additional for $\omega \in R$; $\omega^2 = \frac{\tau^2}{2\cdot(L\cdot C)^2} \pm \frac{1}{2}\cdot\sqrt{\frac{\tau^4}{(L\cdot C)^4} + 4\cdot\frac{1}{(L\cdot C)^2}}$ (162)

And there are two options: first always exist

$$\frac{\tau^2}{2\cdot(L\cdot C)^2} + \frac{1}{2}\cdot\sqrt{\frac{\tau^4}{(L\cdot C)^4} + 4\cdot\frac{1}{(L\cdot C)^2}} > 0 \tag{163}$$

Second $\frac{\tau^2}{2\cdot(L\cdot C)^2} - \frac{1}{2}\cdot\sqrt{\frac{\tau^4}{(L\cdot C)^4} + 4\cdot\frac{1}{(L\cdot C)^2}} < 0$ (164)

$$\omega^2 = \frac{1}{2}\cdot\frac{1}{L\cdot C}\cdot\{\frac{\tau^2}{L\cdot C} \pm \sqrt{\frac{\tau^4}{(L\cdot C)^2} + 4\}} \tag{165}$$

$\sqrt{\frac{\tau^4}{(L\cdot C)^2} + 4} > \frac{\tau^2}{L\cdot C}$, Not exist and always negative for any Cylindrical RF network antennas overall parameters values. We choose only the (+) option (first).

Writing $P(\lambda) = P_R(\lambda) + i\cdot P_I(\lambda)$ and $Q(\lambda) = Q_R(\lambda) + i\cdot Q_I(\lambda)$ (166)

and inserting $\lambda = i\cdot\omega$ Into Cylindrical RF network antennas characteristic equation, ω must satisfy the following:

$$\sin\omega\cdot\tau = g(\omega) = \frac{-P_R(i\cdot\omega)\cdot Q_I(i\cdot\omega) + P_I(i\cdot\omega)\cdot Q_R(i\cdot\omega)}{|Q(i\cdot\omega)|^2} \tag{167}$$

$$\cos\omega\cdot\tau = h(\omega) = -\frac{P_R(i\cdot\omega)\cdot Q_R(i\cdot\omega) + P_I(i\cdot\omega)\cdot Q_I(i\cdot\omega)}{|Q(i\cdot\omega)|^2} \tag{168}$$

Where $|Q(i\cdot\omega)|^2 \neq 0$ in view of requirement (a) above, and $(g,h) \in R$. Furthermore, it follows above $\sin\omega\cdot\tau$ and $\cos\omega\cdot\tau$ equations that, by squaring and adding the sides, ω Must be a positive root of $F(\omega) = |P(i\cdot\omega)|^2 - |Q(i\cdot\omega)|^2 = 0$. (169)

Note that $F(\omega)$ is dependent of τ . Now it is important to

notice that if $\tau \notin I$ (assume that $I \subseteq R_{+0}$ is the set where $\omega(\tau)$ is a positive root of $F(\omega)$ and for $\tau \notin I$, $\omega(\tau)$ is not define. Then for all τ in I $\omega(\tau)$ is satisfies that $F(\omega,\tau) = 0$

Then there are positive $\omega(\tau)$ solutions of $F(\omega,\tau)=0$, and we analyze stability switches. For any $\tau\in I$ where $\omega(\tau)$ is a positive solution of $F(\omega,\tau)=0$, we can define the angle $\theta(\tau)\in[0,2\cdot\pi]$ as the solution of

$$\sin\theta(\tau)=\frac{-P_R(i\cdot\omega)\cdot Q_I(i\cdot\omega)+P_I(i\cdot\omega)\cdot Q_R(i\cdot\omega)}{|Q(i\cdot\omega)|^2} \tag{170}$$

$$\cos\theta(\tau)=-\frac{P_R(i\cdot\omega)\cdot Q_R(i\cdot\omega)+P_I(i\cdot\omega)\cdot Q_I(i\cdot\omega)}{|Q(i\cdot\omega)|^2} \tag{171}$$

And the relation between the argument $\theta(\tau)$ and $\omega(\tau)\cdot\tau$ for $\tau\in I$ must be $\omega(\tau)\cdot\tau=\theta(\tau)+n\cdot 2\cdot\pi\ \ \forall\ \ n\in\mathbb{N}_0$. Hence we can define the maps $\tau_n: I\to R_{+0}$ given by

$\tau_n(\tau)=\dfrac{\theta(\tau)+n\cdot 2\cdot\pi}{\omega(\tau)}$; $n\in\mathbb{N}_0,\tau\in I$. Let us introduce the functions $I\to R$;

$S_n(\tau)=\tau-\tau_n(\tau),\ \tau\in I,\ n\in\mathbb{N}_0$. That are continuous and differentiable in τ. In the following, the subscripts $\lambda,\ \omega,\ C,\ L$ and Cylindrical RF network antennas parameters $(L,C,\tau\ etc.,)$ indicate the corresponding partial derivatives. Let us first concentrate on $\wedge(x)$, remember in $\lambda(L,C,\tau,\ etc.,)$ and $\omega(L,C,\tau,\ etc.,)$, and keeping all parameters except one (x) and τ. The derivation closely follows that in reference [BK]. Differentiating Cylindrical RF network antennas characteristic equation $P(\lambda)+Q(\lambda)\cdot e^{-\lambda\cdot\tau}=0$ with respect to specific parameter (x), and inverting the derivative, for convenience, one calculates: Remark: $x=L,C,\tau,\ etc.,$

$$\begin{aligned}&\left(\frac{\partial\lambda}{\partial x}\right)^{-1}\\&=\frac{-P_\lambda(\lambda,x)\cdot Q(\lambda,x)+Q_\lambda(\lambda,x)\cdot P(\lambda,x)-\tau\cdot P(\lambda,x)\cdot Q(\lambda,x)}{P_x(\lambda,x)\cdot Q(\lambda,x)-Q_x(\lambda,x)\cdot P(\lambda,x)}\end{aligned} \tag{172}$$

Where $P_\lambda=\dfrac{\partial P}{\partial\lambda}$,.... etc., Substituting $\lambda=i\cdot\omega$, and Bearing i $\overline{P(-i\cdot\omega)}=P(i\cdot\omega)$,

$$\overline{Q(-i\cdot\omega)}=Q(i\cdot\omega) \tag{173}$$

Then $i\cdot P_\lambda(i\cdot\omega)=P_\omega(i\cdot\omega)$ and $i\cdot Q_\lambda(i\cdot\omega)=Q_\omega(i\cdot\omega)$ and that on the surface $|P(i\cdot\omega)|^2=|Q(i\cdot\omega)|^2$, one obtains

$$\begin{aligned}&\left(\frac{\partial\lambda}{\partial x}\right)^{-1}|_{\lambda=i\cdot\omega}\\&=\left(\frac{i\cdot P_\omega(i\cdot\omega,x)\cdot\overline{P(i\cdot\omega,x)}+i\cdot Q_\lambda(i\cdot\omega,x)\cdot\overline{Q(\lambda,x)}-\tau\cdot|P(i\cdot\omega,x)|^2}{P_x(i\cdot\omega,x)\cdot\overline{P(i\cdot\omega,x)}-Q_x(i\cdot\omega,x)\cdot\overline{Q(i\cdot\omega,x)}}\right)\end{aligned} \tag{174}$$

Upon separating into real and imaginary parts, with

$$P=P_R+i\cdot P_I\ ;\ Q=Q_R+i\cdot Q_I\ ;\ P_\omega=P_{R\omega}+i\cdot P_{I\omega} \tag{175}$$

$$Q_\omega = Q_{R\omega} + i \cdot Q_{I\omega} \; ; \; P_x = P_{Rx} + i \cdot P_{Ix} \; ; \; Q_x = Q_{Rx} + i \cdot Q_{Ix} \tag{176}$$

$P^2 = P_R^2 + P_I^2$. When (x) can be any Cylindrical RF network antennas parameters L, C, And time delay τ etc,. Where for convenience, we have dropped the arguments $(i \cdot \omega, x)$, and where

$$F_\omega = 2 \cdot [(P_{R\omega} \cdot P_R + P_{I\omega} \cdot P_I) - (Q_{R\omega} \cdot Q_R + Q_{I\omega} \cdot Q_I)] \tag{177}$$

$$F_x = 2 \cdot [(P_{Rx} \cdot P_R + P_{Ix} \cdot P_I) - (Q_{Rx} \cdot Q_R + Q_{Ix} \cdot Q_I)] \tag{178}$$

$\omega_x = -F_x / F_\omega$. We define U and V:

$$U = (P_R \cdot P_{I\omega} - P_I \cdot P_{R\omega}) - (Q_R \cdot Q_{I\omega} - Q_I \cdot Q_{R\omega}) \tag{179}$$

$$V = (P_R \cdot P_{Ix} - P_I \cdot P_{Rx}) - (Q_R \cdot Q_{Ix} - Q_I \cdot Q_{Rx}) \tag{180}$$

We choose our specific parameter as time delay x = τ.

$$V = \frac{\omega}{L^2 \cdot C^2} \; ; U = \frac{\tau}{L^2 \cdot C^2} \; ; P^2 = \omega^4 \; ; F_\tau = \frac{-\omega^2 \cdot 2 \cdot \tau}{(L \cdot C)^2} \tag{181}$$

$$\begin{aligned} &P_R(\omega, \tau) = -\omega^2 \; ; \; P_I(\omega, \tau) = 0 \\ &Q_I(\omega, \tau) = -\frac{\omega \cdot \tau}{L \cdot C} \; ; \; Q_R(\omega, \tau) = \frac{1}{L \cdot C} \end{aligned} \tag{182}$$

$$P_{I\tau} = 0 \; ; P_{R\tau} = 0 \; ; Q_{R\tau} = 0 \; ; Q_{I\tau} = -\frac{\omega}{L \cdot C} \; \Rightarrow V \neq 0 \tag{183}$$

$$\begin{aligned} &\frac{\partial F}{\partial \omega} = F_\omega = 4 \cdot \omega^3 - 2 \cdot \omega \cdot \frac{\tau^2}{(L \cdot C)^2} \\ &\frac{\partial F}{\partial \omega} = 2 \cdot \omega \cdot [2 \cdot \omega^2 - \frac{\tau^2}{(L \cdot C)^2}] \; ; F(\omega, \tau) = 0 \end{aligned} \tag{184}$$

And differentiating with respect to τ and we get

$$F_\omega \cdot \frac{\partial \omega}{\partial \tau} + F_\tau = 0 \; ; \; \tau \in I \Rightarrow \omega_\tau = \frac{\partial \omega}{\partial \tau} = -\frac{F_\tau}{F_\omega} \tag{185}$$

$$\frac{\partial \omega}{\partial \tau} = \frac{\frac{\omega \cdot \tau}{(L \cdot C)^2}}{[2 \cdot \omega^2 - \frac{\tau^2}{(L \cdot C)^2}]} \; ; \; \wedge^{-1}(\tau) = (\frac{\partial \operatorname{Re} \lambda}{\partial \tau})_{\lambda = i \cdot \omega} \tag{186}$$

$$\omega_\tau = \frac{\partial \omega}{\partial \tau} = \frac{\omega \cdot \tau}{[2 \cdot \omega^2 \cdot (L \cdot C)^2 - \tau^2]} \tag{187}$$

$$\wedge^{-1}(\tau)=\mathrm{Re}\{\frac{-2\cdot[U+\tau\cdot|P|^2]+i\cdot F_\omega}{F_\tau+i\cdot 2\cdot[V+\omega\cdot|P|^2]}\}$$
$$=\mathrm{Re}\{\frac{-\tau\cdot[\frac{1}{L^2\cdot C^2}+\omega^4]+i\cdot\omega\cdot[2\cdot\omega^2-\frac{\tau^2}{(L\cdot C)^2}]}{\frac{-\omega^2\cdot\tau}{(L\cdot C)^2}+i\cdot\omega\cdot[\frac{1}{L^2\cdot C^2}+\omega^4]}\} \quad (188)$$

$$sign\{\wedge^{-1}(\tau)\}=sign\{(\frac{\partial\,\mathrm{Re}\,\lambda}{\partial\tau})_{\lambda=i\cdot\omega}\} \quad (189)$$

$$sign\{\wedge^{-1}(\tau)\}=sign\{F_\omega\}\cdot sign\{\tau\cdot\frac{\partial\omega}{\partial\tau}+\omega+\frac{U\cdot\frac{\partial\omega}{\partial\tau}+V}{|P|^2}\} \quad (190)$$

$$sign\{\wedge^{-1}(\tau)\}=sign\{2\cdot\omega\cdot[2\cdot\omega^2-\frac{\tau^2}{(L\cdot C)^2}]\}$$
$$\cdot sign\{\tau\cdot[\frac{\frac{\omega\cdot\tau}{(L\cdot C)^2}}{[2\cdot\omega^2-\frac{\tau^2}{(L\cdot C)^2}]}+\omega$$
$$+\frac{\frac{\tau}{L^2\cdot C^2}\cdot[\frac{\frac{\omega\cdot\tau}{(L\cdot C)^2}}{[2\cdot\omega^2-\frac{\tau^2}{(L\cdot C)^2}]}+\frac{\omega}{L^2\cdot C^2}}{\omega^4}\}] \quad (191)$$

We define new variables: ψ_1,ψ_2,ψ_3

$$\psi_1(\omega,\tau,L,C)=2\cdot\omega\cdot[2\cdot\omega^2-\frac{\tau^2}{(L\cdot C)^2}] \quad (192)$$

$$\psi_2(\omega,\tau,L,C)=\tau\cdot[\frac{\frac{\omega\cdot\tau}{(L\cdot C)^2}}{[2\cdot\omega^2-\frac{\tau^2}{(L\cdot C)^2}]}] \quad (193)$$

$$\psi_3(\omega,\tau,L,C)=\frac{\frac{\tau}{L^2\cdot C^2}\cdot[\frac{\frac{\omega\cdot\tau}{(L\cdot C)^2}}{[2\cdot\omega^2-\frac{\tau^2}{(L\cdot C)^2}]}+\frac{\omega}{L^2\cdot C^2}}{\omega^4} \quad (194)$$

$$sign\{\wedge^{-1}(\tau)\}=sign[\psi_1]\cdot sign[\psi_2+\omega+\psi_3]$$

We check the sign of $\wedge^{-1}(\tau)$ according the following rule:

$sign[F_\omega]$	$sign[\frac{V+\omega_\tau \cdot U}{P^2}+\omega+\omega_\tau \cdot \tau]$	$sign[\wedge^{-1}(\tau)]$
+/-	+/-	+
+/-	-/+	-

Table 2. Cylindrical RF network antennas system stability switching criteria.

If sign $[\Lambda^{-1}(\tau)] > 0$ then the crossing proceeds from (-) to (+) respectively (stable to unstable). If sign$[\Lambda^{-1}(\tau)] < 0$ then the crossing proceeds from (+) to (-) respectively (unstable to stable). Anyway the stability switching can occur only for specific ω, τ. Since it is a very complex function, we recommend to solve it numerically rather than analytic. We plot the stability switch diagram based on different delay values of our Cylindrical RF network antennas system.

$$D(\lambda,\tau_1=\tau_2=\tau)=\lambda^2+\frac{1}{L\cdot C}\cdot e^{-\lambda\cdot\tau}-\lambda\cdot\frac{\tau}{L\cdot C}\cdot e^{-\lambda\cdot\tau} \tag{195}$$

Taylor expansion: $e^{-\lambda\tau}\approx 1-\lambda\cdot\tau+\frac{\lambda^2\cdot\tau^2}{2}$ since we need

n > m [BK] analysis we choose $e^{-\lambda\tau}\approx 1-\lambda\cdot\tau$ then we get

Our Cylindrical RF network antennas system second order characteristic equation:

$$\begin{aligned}&D(\lambda,\tau)=\lambda^2+a(\tau)\cdot\lambda+b(\tau)\cdot\lambda\cdot e^{-\lambda\cdot\tau}\\&+c(\tau)+d(\tau)\cdot e^{-\lambda\cdot\tau}\end{aligned} \tag{196}$$

$$a(\tau)=0\ ;\ b(\tau)=-\frac{\tau}{L\cdot C}\ ;\ c(\tau)=0\ ;\ d(\tau)=\frac{1}{L\cdot C} \tag{197}$$

$$\begin{aligned}&F(\omega,\tau)=|P(i\cdot\omega,\tau)|^2-|Q(i\cdot\omega,\tau)|^2=(c-\omega^2)^2\\&+\omega^2\cdot a^2-(\omega^2\cdot b^2+d^2)\end{aligned} \tag{198}$$

$$F(\omega,\tau)=\omega^4-\omega^2\cdot\frac{\tau^2}{(L\cdot C)^2}-\frac{1}{(L\cdot C)^2} \tag{199}$$

Hence

$$F(\omega,\tau)=0 \text{ implies } \omega^4-\omega^2\cdot\frac{\tau^2}{(L\cdot C)^2}-\frac{1}{(L\cdot C)^2}=0 \tag{200}$$

And its roots are given by

$$\omega_+^2=\frac{1}{2}\cdot\{(b^2+2\cdot c-a^2)+\sqrt{\Delta}\}=\frac{1}{2}\cdot\{\sqrt{\Delta}+\frac{\tau^2}{(L\cdot C)^2}\} \tag{201}$$

$$\omega_-^2=\frac{1}{2}\cdot\{(b^2+2\cdot c-a^2)-\sqrt{\Delta}\}=\frac{1}{2}\cdot\{-\sqrt{\Delta}+\frac{\tau^2}{(L\cdot C)^2}\} \tag{202}$$

$$\Delta=(b^2+2\cdot c-a^2)-4\cdot(c^2-d^2)=\frac{\tau^2+4}{(L\cdot C)^2} \tag{203}$$

Therefore the following holds:

$$2\cdot\omega_{+/-}^2-(b^2+2\cdot c-a^2)=\pm\sqrt{\Delta} \tag{204}$$

$$\sin\theta(\tau)=\frac{-P_R(i\cdot\omega,\tau)\cdot Q_I(i\cdot\omega,\tau)+P_I(i\cdot\omega,\tau)\cdot Q_R(i\cdot\omega,\tau)}{|Q(i\cdot\omega,\tau)|^2} \tag{205}$$

$$\cos\theta(\tau)=-\frac{P_R(i\cdot\omega,\tau)\cdot Q_R(i\cdot\omega,\tau)+P_I(i\cdot\omega,\tau)\cdot Q_I(i\cdot\omega,\tau)}{|Q(i\cdot\omega,\tau)|^2} \tag{206}$$

$$\sin\theta(\tau)=\frac{-(c-\omega^2)\cdot\omega\cdot b+\omega\cdot a\cdot d}{\omega^2\cdot b^2+d^2}=\frac{-\omega^3\cdot\tau\cdot L\cdot C}{(\omega^2\cdot\tau^2+1)} \tag{207}$$

$$\cos\theta(\tau)=-\frac{(c-\omega^2)\cdot d+\omega^2\cdot a\cdot b}{\omega^2\cdot b^2+d^2}=\frac{\omega^2\cdot L\cdot C}{(\omega^2\cdot\tau^2+1)} \tag{208}$$

We consider Cylindrical RF antenna which mounted outside a Pyrex glass tube of diameter 32cm and length 50cm. The RF antenna consists of 16 copper (Cu) legs equally spaced by 6.7cm interconnected with capacitors of 2.47nF. Copper leg diameter is equal to 1mm and length 30cm=300mm (<Pyrex glass tube length, 50cm). We consider for Copper (Cu), relative permeability is one. f=10MHz is the typical testing frequency for cylindrical (birdcage) antenna. L – Inductance (nH), l – length of copper leg (mm), d – diameter of copper leg, f – testing frequency. l>100·d (300mm>100·1mm), $d^2\cdot f>1mm^2\cdot MHz$ ($1mm^2\cdot 10MHz>1mm^2\cdot MHz$). L=365.4nH. $L=\frac{1}{5}\cdot l\cdot[\ln(\frac{4\cdot l}{d}-1]=365.4nH$. For our stability switching analysis we choose typical Cylindrical RF network antennas parameters values (as calculated):

$C=2.47nF$; $L=365.4nH$; $R_p=100ohm$ then $\frac{1}{L\cdot C}=0.00110798\cdot 10^{18}$. We find those ω, τ values which fulfill $F(\omega,\tau)=0$. We ignore negative, complex, and imaginary values of ω for specific τ values. The below table gives the list. Remark: we know $F(\omega,\tau)=0$ implies it roots $\omega_i(\tau)$ and finding those delays values τ which ω_i is feasible. There are τ values , which ω_i are complex or imaginary numbered, then unable to analyze stability [6] [7].

6 CYLINDRICAL RF NETWORK ANTENNAS SYSTEM STABILITY ANALYSIS UNDER DELAYED VARIABLES IN TIME $\tau_1=\tau$; $\tau_2=\tau$, RESULTS

We find those ω, τ values which fulfill $F(\omega,\tau)=0$. We ignore negative, complex, and imaginary values of ω for specific τ values. $\tau\in[0.001..10]$ and we can be express by 3D function $F(\omega,\tau)=0$.

$$F(\omega,\tau)=\omega^4-\omega^2\cdot\frac{\tau^2}{(L\cdot C)^2}-\frac{1}{(L\cdot C)^2} \tag{209}$$

$$F(\omega,\tau) = |P(i\cdot\omega,\tau)|^2 - |Q(i\cdot\omega,\tau)|^2$$
$$= \Phi_0 + \Phi_2\cdot\omega^2 + \Phi_4\cdot\omega^4 = \sum_{k=0}^{2}\Phi_{2\cdot k}\cdot\omega^{2\cdot k} \quad (210)$$

$$\Phi_0 = -\frac{1}{(L\cdot C)^2}\ ;\ \Phi_2 = -\frac{\tau^2}{(L\cdot C)^2}\ ;\ \Phi_4 = 1 \quad (211)$$

Hence $F(\omega,\tau)=0$ implies $\sum_{k=0}^{4}\Phi_{2\cdot k}\cdot\omega^{2\cdot k} = 0$ (212)

$\Phi_j \rightarrow Phi_j$. Running MATLAB script for τ values ($\tau \in [0.001..10]$) gives the following results:

MATLAB script: Tau=0.001;C=2.47*1e-9;L=365.4*1e-9;Phi0=-1/(C*L*C*L); Phi2=-(Tau*Tau)/(C*L*C*L); Phi4=1;p=[Phi4 0 Phi2 0 Phi0];r=roots(p)

τ	τ=0.01sec	τ=0.001sec
ω_1	1.0e+013 *	1.0e+012 *
ω_2	-1.1080	-1.1080
ω_3	1.1080	1.1080
ω_4	0.0000 + 0.0000i	-0.0000 + 0.0000i
ω_5	0.0000 - 0.0000i	-0.0000 - 0.0000i

Table 3a. Cylindrical RF network antennas system roots $\omega_i(\tau)$

τ	τ=1sec	τ=0.1sec
ω_1	1.0e+015 *	1.0e+014 *
ω_2	-1.1080	-1.1080
ω_3	1.1080	1.1080
ω_4	-0.0000 + 0.0000i	0.0000 + 0.0000i
ω_5	-0.0000 - 0.0000i	0.0000 - 0.0000i

Table 3b. Cylindrical RF network antennas system roots $\omega_i(\tau)$

τ	τ=3sec	τ=2sec
ω_1	1.0e+015 *	1.0e+015 *
ω_2	3.3240	-2.2160
ω_3	-3.3240	2.2160
ω_4	0 + 0.0000i	-0.0000 + 0.0000i
ω_5	0 - 0.0000i	-0.0000 - 0.0000i

Table 3c. Cylindrical RF network antennas system roots $\omega_i(\tau)$

τ	τ=5sec	τ=4sec
ω_1	1.0e+015 *	1.0e+015 *
ω_2	-5.5399	4.4319
ω_3	5.5399	-4.4319
ω_4	0.0000 + 0.0000i	0 + 0.0000i
ω_5	0.0000 - 0.0000i	0 - 0.0000i

Table 3d. Cylindrical RF network antennas system roots $\omega_i(\tau)$

τ	τ=7sec	τ=6sec
ω_1	1.0e+015 *	1.0e+015 *
ω_2	-7.7559	6.6479
ω_3	7.7559	-6.6479
ω_4	0.0000 + 0.0000i	0 + 0.0000i
ω_5	0.0000 - 0.0000i	0 - 0.0000i

Table 3e. Cylindrical RF network antennas system roots $\omega_i(\tau)$

τ	τ=9sec	τ=8sec
ω_1	1.0e+015 *	1.0e+015 *
ω_2	9.9719	8.8639
ω_3	-9.9719	-8.8639
ω_4	0 + 0.0000i	0 + 0.0000i
ω_5	0 - 0.0000i	0 - 0.0000i

Table 3f. Cylindrical RF network antennas system roots $\omega_i(\tau)$

Table 3g. Cylindrical RF network antennas system roots $\omega_i(\tau)$

τ	τ=0sec	τ=10sec
ω_1	1.0e+007 *	1.0e+016 *
ω_2	-3.3286	-1.1080
ω_3	-0.0000 + 3.3286i	1.1080
ω_4	-0.0000 - 3.3286i	-0.0000 + 0.0000i
ω_5	3.3286	-0.0000 - 0.0000i

We can summery our $\omega_i(\tau)$ results for $\omega_i(\tau) > 0$ and real number (ignore complex, negative, and imaginary values). We exclude from our table the high and real $\omega_i(\tau)$ values (1.0e+007 *, 1.0e+012 *,…,1.0e+016 *) and add results for τ=15sec and τ=20sec.

τ[sec]	ω	$\sin(\omega\cdot\tau) = \frac{-\omega^3\cdot\tau\cdot L\cdot C}{(\omega^2\cdot\tau^2+1)}$	$\cos(\omega\cdot\tau) = \frac{\omega^2\cdot L\cdot C}{(\omega^2\cdot\tau^2+1)}$
0	3.3286	0 =0	1≠9.9e-15
0.001..1	1.1080	-1.22e-18… -5.51e-16	1.108e-15… 4.973e-16
2	2.2160	-9.5e-16	2.1e-16
3	3.3240	-9.9e-16	9.9e-17
4	4.4319	-9.9e-16	5.62e-17
5	5.5399	-9.9e-16	3.6e-17
6	6.6479	-9.99-16	2.5055e-17
7	7.7559	-9.9966e-16	1.8413e-17
8	8.8639	-9.9980e-16	1.4099e-17
9	9.9719	-9.9988e-16	1.1141e-17
10	1.1080	-9.9193e-17	8.9525e-18
15	1.6620	-9.9841e-17	4.0048e-18
20	2.2160	-9.9950e-17	2.2552e-18

Table 4. Cylindrical RF network antennas system positive and real roots $\omega_i(\tau)$ values and $\sin(\omega\cdot\tau),\ \cos(\omega\cdot\tau)$ values.

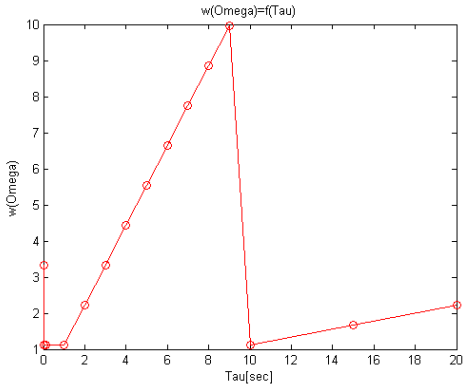


Figure5. Cylindrical RF network F(ω,τ) function for $\tau_1=\tau_2=\tau$

Matlab: plot([0 0.001 0.01 0.1 1 2 3 4 5 6 7 8 9 10 15 20],[3.3286 1.1080 1.1080 1.1080 1.1080 2.2160 3.3240 4.4319 5.5399 6.6479 7.7559 8.8639 9.9719 1.1080 1.6620 2.2160],'-or'). We plot 3D function $F(\omega,\tau)=0$. τ:0→10 ; ω:0→100. We define additional MATLAB script parameters ω→w, τ→t.

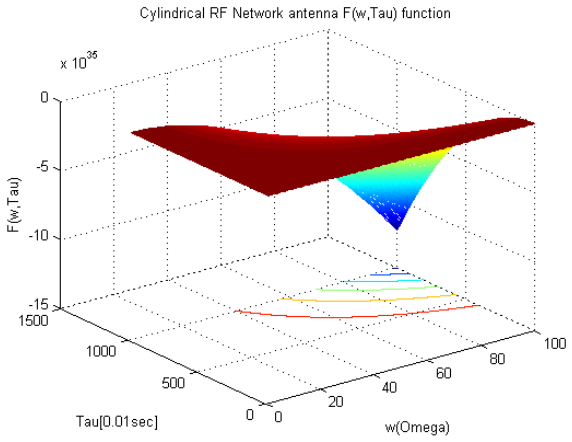


Figure 6. Cylindrical RF network F(ω,τ) function for $\tau_1=\tau_2=\tau$

Matlab:[w,t]=meshgrid(1:1:100,0:0.01:10);C=2.47*1e-9;L=365.4*1e-9;f=w.*w.*w.*w-w.*w.*(t.*t)/(C*L*C*L)-1/(C*L*C*L);meshc(f); % $\omega \to w, \tau \to t$

We get two possible real values for ω which fulfil $F(\omega,\tau)=0$

$F(\omega=3.3286 \text{ or } \omega=1.1080 \text{ ..... or } \omega=2.2160,\tau)=0$; $\tau \in [0.001..10]$ Next is to find those ω, τ values which fulfil $\sin\theta(\tau)=...$

$$\sin(\omega\cdot\tau)=\frac{-P_R\cdot Q_I+P_I\cdot Q_R}{|Q|^2} \text{ And } \cos\theta(\tau)=... \tag{213}$$

$$\cos(\omega\cdot\tau)=-\frac{(P_R\cdot Q_R+P_I\cdot Q_I)}{|Q|^2} \; ; \; |Q|^2=Q_R^2+Q_I^2 \cdot \tag{214}$$

$$\sin(\omega\cdot\tau)=\frac{-\omega^3\cdot\tau\cdot L\cdot C}{(\omega^2\cdot\tau^2+1)} \; ; \; \cos(\omega\cdot\tau)=\frac{\omega^2\cdot L\cdot C}{(\omega^2\cdot\tau^2+1)} \tag{215}$$

$\frac{-\omega^3\cdot\tau\cdot L\cdot C}{(\omega^2\cdot\tau^2+1)}<0$ & $\frac{\omega^2\cdot L\cdot C}{(\omega^2\cdot\tau^2+1)}>0$ then $\sin(\omega\cdot\tau)<0$ and $\cos(\omega\cdot\tau)>0$; $2\cdot\pi>\omega\cdot\tau>\frac{\pi}{2}\cdot 3$. We plot the stability switch diagram based on different delay values of our Cylindrical RF network antennas system.

$$\wedge^{-1}(\tau)=(\frac{\partial \operatorname{Re}\lambda}{\partial\tau})_{\lambda=i\cdot\omega}=\operatorname{Re}\{\frac{-2\cdot[U+\tau\cdot|P|^2]+i\cdot F_\omega}{F_\tau+i\cdot 2\cdot[V+\omega\cdot|P|^2]}\} \tag{216}$$

$$\wedge^{-1}(\tau)=(\frac{\partial \operatorname{Re}\lambda}{\partial\tau})_{\lambda=i\cdot\omega}=\frac{2\cdot\{F_\omega\cdot(V+\omega\cdot P^2)-F_\tau\cdot(U+\tau\cdot P^2)\}}{F_\tau^2+4\cdot(V+\omega\cdot P^2)^2} \tag{217}$$

$sign\{\wedge^{-1}(\tau)\}=sign[\psi_1]\cdot sign[\psi_2+\omega+\psi_3]$. We define the following new functions: $g_1=\psi_1$; $g_2=\psi_2+\omega+\psi_3$ (218)

$$sign\{\wedge^{-1}(\tau)\}=sign[g_1]\cdot sign[g_2] \tag{219}$$

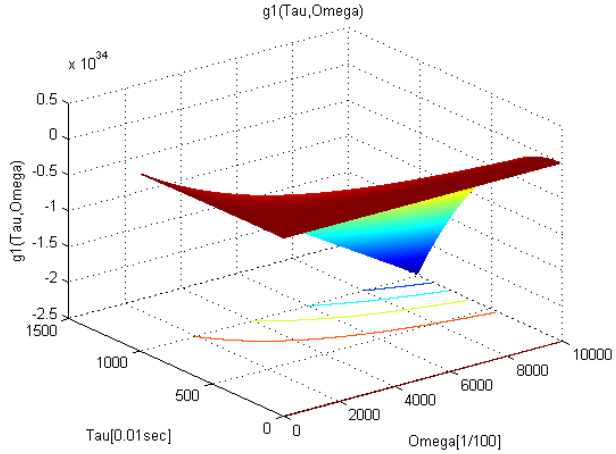


Figure 7. Cylindrical RF network $g_1(\tau,\omega)$ function for $\tau_1=\tau_2=\tau$.

Matlab:[w,t]=meshgrid(1:.01:100,0:0.01:10);C=2.47*1e-9; L=365.4*1e-9;f=2*w.*(2*w.*w-(t.*t./(C*L*C*L)));meshc(f)

% $\omega \to w, \tau \to t$

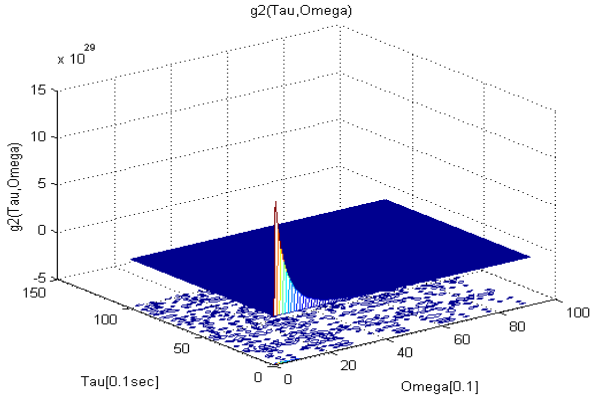


Figure 8. Cylindrical RF network $g_2(\tau,\omega)$ function for $\tau_1=\tau_2=\tau$.

Matlab:[w,t]=meshgrid(1:.1:10,0:0.1:10);C=2.47*1e-9; L=365.4*1e-9;m=w.*t./(2*w.*w.*(L*C*L*C)-t.*t); f=t.*m+w+(m.*t./(L*C*L*C)+w./(L*C*L*C))./(w.*w.*w.*w);meshc(f) % $\omega \to w, \tau \to t$

$$g(Tau) = g_1(Tau)\cdot g_2(Tau) = \wedge^{-1}(\tau) = \left(\frac{\partial \operatorname{Re}\lambda}{\partial \tau}\right)_{\lambda=i\cdot\omega} \tag{220}$$

The stability switch occur only on those delay values (τ) which fit the equation : $\tau = \frac{\theta_+(\tau)}{\omega_+(\tau)}$ and $\theta_+(\tau)$ is the solution

Of $\sin\theta(\tau) = \frac{-\omega^3\cdot\tau\cdot L\cdot C}{(\omega^2\cdot\tau^2+1)}$; $\cos\theta(\tau) = \frac{\omega^2\cdot L\cdot C}{(\omega^2\cdot\tau^2+1)}$ (221)

When $\omega = \omega_+(\tau)$ if only ω_+ is feasible. Additionally When all Cylindrical RF network antennas parameters are known and the stability switch due to various time delay values τ is describe in the below expression (Appendix E):

$$sign\{\wedge^{-1}(\tau)\} = sign\{F_\omega(\omega(\tau),\tau)\} \cdot sign\{\tau \cdot \omega_\tau(\omega(\tau)) + \omega(\tau) + \frac{U(\omega(\tau)) \cdot \omega_\tau(\omega(\tau)) + V(\omega(\tau))}{|P(\omega(\tau))|^2}\} \quad (222)$$

Remark: we know $F(\omega,\tau) = 0$ implies it roots $\omega_i(\tau)$ and finding those delays values τ which ω_i is feasible. There are τ values which ω_i is complex or imaginary number, then unable to analyse stability [4] [5].

7 Discussion

In this paper, we consider Cylindrical RF network antennas system. Due to RF antenna copper leg parasitic effect we get copper leg's current and current derivative with delay τ1-k and τ2-k (k is leg number index, k=1,…,16).. Those delays causes to stability switching for our Cylindrical RF network antennas. We draw our Cylindrical RF network antennas equivalent circuit and get system differential equations. Our variables are Y, X which are function of RF antenna copper leg's current and current derivative. Our system dynamic behavior is dependent on circuit overall parameters and parasitic delay in time. We keep all circuit parameters fix and change, parasitic delay over various values $\tau \in [0.001..10]$. Our analysis results extend that of in the way that it deals with stability switching for different delay values. This implies that our system behavior of the circuit cannot inspect by short analysis and we must study the full system. Several very important issues and possibilities were left out of the present paper. One possibility is the stability switching by circuit parameters. Every circuit's parameter variation can change our system dynamic and stability behavior. This case can be solved by the same methods combined with alternative and more technical hypotheses. Moreover, numerical simulations for the Cylindrical RF network antennas model studied in references suggest that this result can be extended to enhance models with more general functions. Still another extension of our results would be to treat the case of delayed Cylindrical RF network antennas leg's higher derivative degree of current. It would be extremely desirable to confirm these cases by mathematical proofs.

8 Conclusion

Cylindrical RF network antennas system is characterized by parasitic effects which can influence Cylindrical RF network antennas system stability in time. There are two main Cylindrical RF network antennas variables which are affected by antenna legs parasitic effects, Y and X functions of antenna leg's currents and currents derivatives respectively. Each Cylindrical RF network antennas system variable under parasitic effects is characterized by time delay respectively. The two time delays are not the same, but can be categorized to some sub cases due to antenna leg parasitic behavior.

The first case we analyze is when there is delay in Cylindrical RF network antennas leg's current and no delay in antennas leg's current derivative or opposite. The second case we analyze is when there is delay both in Cylindrical RF network antennas leg's current and current time derivative [4] [5]. For simplicity of our analysis we consider in the second case all delays are the same (there is a difference but it is

neglected in our analysis). In each case we derive the related characteristic equation. The characteristic equation is dependent on Cylindrical RF network antennas system overall parameters and parasitic time delay. Upon mathematics manipulation and [BK] theorems and definitions we derive the expression which gives us a clear picture on Cylindrical RF network antennas map. The stability map gives all possible options for stability segments, each segment belongs to different time delay value segment. Cylindrical RF network antennas system's stability analysis can be influenced either by system overall parameter values. We left this analysis and do not discuss it in the current article.

Appendix A : Lemma 1.1

Assume that $\omega(\tau)$ is a positive and real root of $F(\omega,\tau)=0$

Defined for $\tau \in I$, which is continuous and differentiable. Assume further that if $\lambda = i\cdot\omega$, $\omega \in R$, then $P_n(i\cdot\omega,\tau)+Q_n(i\cdot\omega,\tau)\neq 0,\ \tau\in \mathrm{R}$ hold true. Then the functions $S_n(\tau),\ n\in N_0$, are continuous and differentiable on I.

Appendix B : Theorem 1.2

Assume that $\omega(\tau)$ is a positive real root of $F(\omega,\tau)=0$ defined for $\tau\in I,\ I\subseteq R_{+0}$, and at some $\tau^*\in I$, $S_n(\tau^*)=0$

For some $n\in N_0$ then a pair of simple conjugate pure imaginary roots $\lambda_+(\tau^*)=i\cdot\omega(\tau^*),\ \lambda_-(\tau^*)=-i\cdot\omega(\tau^*)$ of $D(\lambda,\tau)=0$ exist at $\tau=\tau^*$ which crosses the imaginary axis from left to right if $\delta(\tau^*)>0$ and cross the imaginary axis from right to left if $\delta(\tau^*)<0$ where

$$\delta(\tau^*)=sign\{\frac{d\,\mathrm{Re}\,\lambda}{d\tau}|_{\lambda=i\omega(\tau^*)}\}= sign\{F_\omega(\omega(\tau^*),\tau^*)\}\cdot sign\{\frac{dS_n(\tau)}{d\tau}|_{\tau=\tau^*}\} \tag{223}$$

The theorem becomes

$$sign\{\frac{d\,\mathrm{Re}\,\lambda}{d\tau}|_{\lambda=i\omega\pm}\}=sign\{\pm\Delta^{1/2}\}\cdot sign\{\frac{dS_n(\tau)}{d\tau}|_{\tau=\tau^*}\} \tag{224}$$

Appendix C : Theorem 1.3

The characteristic equation: $\tau_1=\tau,\tau_2=0;\ \tau_1=0,\tau_2=\tau$

$$D(\lambda,\tau)=\lambda^2+a(\tau)\cdot\lambda+b(\tau)\cdot\lambda\cdot e^{-\lambda\cdot\tau}+c(\tau)+d(\tau)\cdot e^{-\lambda\cdot\tau} \tag{225}$$

$$D(\lambda,\tau_1,\tau_2)=\lambda^2+\lambda\cdot\frac{1}{C1\cdot R1}\cdot e^{-\lambda\cdot\tau_2}+\frac{1}{C1\cdot f_\#}\cdot e^{-\lambda\cdot(\tau_1+\tau_2)} \tag{226}$$

Has a pair of simple and conjugate pure imaginary roots

$\lambda = \pm\omega(\tau^*)$, $\omega(\tau^*)$ real at $\tau^* \in I$ if $S_n(\tau^*) = \tau^* - \tau_n(\tau^*) = 0$ for some $n \in N_0$. If $\omega(\tau^*) = \omega_+(\tau^*)$, this pair of simple conjugate pure imaginary roots crosses the imaginary axis from left to right if $\delta_+(\tau^*) > 0$ and crosses the imaginary axis from right to left if $\delta_+(\tau^*) < 0$ where

$$\delta_+(\tau^*) = sign\{\frac{d\,\mathrm{Re}\,\lambda}{d\tau}|_{\lambda=i\omega_+(\tau^*)}\} = sign\{\frac{dS_n(\tau)}{d\tau}|_{\tau=\tau^*}\} \tag{227}$$

If $\omega(\tau^*) = \omega_-(\tau^*)$, this pair of simple conjugate pure imaginary roots cross the imaginary axis from left to right if

$\delta_-(\tau^*) > 0$ and crosses the imaginary axis from right to left

If $\delta_-(\tau^*) < 0$ where

$$\delta_-(\tau^*) = sign\{\frac{d\,\mathrm{Re}\,\lambda}{d\tau}|_{\lambda=i\omega_-(\tau^*)}\} = -sign\{\frac{dS_n(\tau)}{d\tau}|_{\tau=\tau^*}\} \tag{228}$$

If $\omega_+(\tau^*) = \omega_-(\tau^*) = \omega(\tau^*)$ then $\Delta(\tau^*) = 0$ and $sign\{\frac{d\,\mathrm{Re}\,\lambda}{d\tau}|_{\lambda=i\omega(\tau^*)}\} = 0$, the same is true when $S_n'(\tau^*) = 0$

The following result can be useful in identifying values of τ

Where stability switches happened.

Appendix D : Theorem 1.4

Assume that for all $\tau \in I$, $\omega(\tau)$ is defined as a solution of

$F(\omega,\tau) = 0$ then $\delta_\pm(\tau) = sign\{\pm\Delta^{1/2}(\tau)\}\cdot signD_\pm(\tau)$

$$\begin{aligned} D_\pm(\tau) &= \omega_\pm^2\cdot[(\omega_\pm^2\cdot b^2 + d^2) + a'\cdot(c-\omega_\pm^2) + b\cdot d' - b'\cdot d - a\cdot c'] \\ &+\omega_\pm\cdot\omega_\pm'\cdot[\tau\cdot(\omega_\pm^2\cdot b^2 + d^2) - b\cdot d + a\cdot(c-\omega_\pm^2) + 2\cdot\omega_\pm^2\cdot a] \end{aligned} \tag{229}$$

$$a' = \frac{da(\tau)}{d\tau}\ ;\ b' = \frac{db(\tau)}{d\tau}\ ;\ c' = \frac{dc(\tau)}{d\tau}\ ;\ d' = \frac{dd(\tau)}{d\tau} \tag{230}$$

Appendix E: We need to approve the following expression

$$\begin{aligned} sign\{\wedge^{-1}(\tau)\} &= sign\{F_\omega(\omega(\tau),\tau)\}\cdot sign\{\tau\cdot\omega_\tau(\omega(\tau)) \\ &+\omega(\tau) + \frac{U(\omega(\tau))\cdot\omega_\tau(\omega(\tau)) + V(\omega(\tau))}{|P(\omega(\tau))|^2}\} \end{aligned} \tag{231}$$

The basic assumption: $\wedge^{-1}(\tau) = (\frac{\partial\,\mathrm{Re}\,\lambda}{\partial\tau})_{\lambda=i\cdot\omega}$ (232)

$$\wedge^{-1}(\tau) = (\frac{\partial\,\mathrm{Re}\,\lambda}{\partial\tau})_{\lambda=i\cdot\omega} = \frac{2\cdot\{F_\omega\cdot(V+\omega\cdot P^2) - F_\tau\cdot(U+\tau\cdot P^2)\}}{F_\tau^2 + 4\cdot(V+\omega\cdot P^2)^2} \tag{233}$$

$sign\{F_\tau^{\,2}+4\cdot(V+\omega\cdot P^2)^2\}>0$ and $\omega_\tau=-\frac{F_\tau}{F_\omega}$ then

$$sign\{\wedge^{-1}(\tau)\}=sign\{(\frac{\partial \operatorname{Re}\lambda}{\partial\tau})_{\lambda=i\cdot\omega}\}$$
$$=sign\{F_\omega\cdot(V+\omega\cdot P^2)-F_\tau\cdot(U+\tau\cdot P^2)\} \quad (234)$$

$$sign\{\wedge^{-1}(\tau)\}=sign\{F_\omega\cdot\{(V+\omega\cdot P^2)-\frac{F_\tau}{F_\omega}\cdot(U+\tau\cdot P^2)\}\} \quad (235)$$

$$sign\{\wedge^{-1}(\tau)\}=sign\{F_\omega\cdot\{(V+\omega\cdot P^2)+\omega_\tau\cdot(U+\tau\cdot P^2)\}\} \quad (236)$$

$$sign\{\wedge^{-1}(\tau)\}=sign\{F_\omega\cdot\{V+\omega_\tau\cdot U+\omega\cdot P^2+\omega_\tau\cdot\tau\cdot P^2\}\} \quad (237)$$

$$sign\{\wedge^{-1}(\tau)\}=sign\{P^2\cdot F_\omega\cdot\{\frac{V+\omega_\tau\cdot U}{P^2}+\omega+\omega_\tau\cdot\tau\}\} \quad (238)$$

$$sign\{\wedge^{-1}(\tau)\}=sign\{P^2\}\cdot sign\{F_\omega\}$$
$$\cdot sign\{\frac{V+\omega_\tau\cdot U}{P^2}+\omega+\omega_\tau\cdot\tau\}\ ;\ sign\{P^2\}>0 \quad (239)$$

$$sign\{\wedge^{-1}(\tau)\}=sign\{F_\omega\}\cdot sign\{\frac{V+\omega_\tau\cdot U}{P^2}+\omega+\omega_\tau\cdot\tau\} \quad (240)$$

REFERENCES

[1]. Ch Hollenstein, Ph Guittienne and A A Howling, Resonant RF network antennas for large-area and large-volume inductively coupled plasma sources, Plasma Sources Sci. Technol. 22 (2013) 055021 (10pp).

[2]. Yuri A. *Kuznetsov, Elelments of Applied Bifurcation Theory*. Applied Mathematical Sciences. Springer 1995.

[3]. Jack K. Hale., Huseyin Kocak. *Dynamics and Bifurcations*.Texts in Applied Mathematics, Vol. 3. Springer-Verlag 1991.

[4]. Steven H. Strogatz, *Nonlinear Dynamics and Chaos*. Westview press. 1994

[5]. Kuang, Y., 1993. Delay Differential Equations with Applications in Population Dynamics. Academic Press, Boston.

[6]. Beretta E., Kuang, Y., 2002. Geometric stability switch criteria in delay differential systems with delay dependent parameters. SIAM J. Math. Anal. 33, 1144-1165.

[7]. Gerard looss., Daniel D.Joseph. Elementary stability and Bifurcation theory. Springer 1980.

[8]. John Guckenheimer., Philip Holmes. Nonlinear oscillations, dynamical systems, and bifurcation of vector fields. Applied Mathematical Sciences 42. Springer 1983.

[9]. Lawrence Perko. Differential equations and dynamics systems, Texts in applied mathematics 7. Springer 1991.

[10]. Kuang Jiaoxun and Cong Yuhao. Stability of numerical methods for delay differential equations, since press USA Inc, 2005.

TNC Transactions on Networks and Communications
SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
VOLUME 2, ISSUE 5
ISSN: 2054 - 7420

N-Cryptographic Multilevel Algorithm for Effective Information Security

Olawale S. Adebayo[a], [b]Morufu Olalere, [c]Amit. Mishra, [d]M. A. Mabayoje and [e]Joel N. Ugwu
[a,b,e] Cyber Security Science, School of Information and Communication Technology, Federal University of Technology, Minna, Niger State, Nigeria, [c]IBB University, Lapai
waleadebayo@futminnna.edu.ng; lerejide@futminna.edu.ng; joeljupiter@yahoo.com

ABSTRACT

Security of information cannot be perfectly realized as both it and its counter technology continue to evolve. In the same vein, using a single cryptographic cipher to realize information secrecy is not enough as it can be broken over time, thus revealing the information in plaintext. Most of the existing cryptographic ciphers possess a minimal level of weakness, which is exploitable over time, but when this algorithmic transformation is used in multiple times, the number of trials, effort, and time required to exploit it become greater. This paper proposes a multilevel algorithm for realizing the security of information using a multiple-cryptographic ciphers process. It presents the possibility of combining n-cryptographic ciphers in a single implementation within the system, permitting n-cryptographic transformation to take place during the encryption and decryption processes.

Keywords— Encryption, Cryptography, n-cryptographic algorithms, Cryptanalysis, Multilevel Algorithm

1 Introduction

The knowledge of cryptography is evolving. Different cryptosystems have been formulated, and people can now make choices of their implementations ranging from symmetric keys to asymmetric ones. The choice depends on the application of the system. However, the knowledge of cryptanalysis has made realizing perfect secrecy a difficult task; people can now make trials of several keys on a particular cipher in order to decrypt encrypted information. The more stringent an algorithm is, the more effort and time required to realize its encrypted information in plaintext. Such efforts can now be automated, thereby reducing the total time of trial to the nearest minimum. As cryptologists continue to discover more algorithms that will ensure better security and frustrate the existing hacking software; hackers continue to rediscover a means of beating the new ones by developing new systems that can combat the new algorithm.

In the past, scientists have found it difficult to ensure an individual can communicate thought, ideas, knowledge, etc. to another without unintended individuals also having access. This difficulty has resulted in many research projects in the area of information security and secrecy, even before the development of electronic computers. The search could be traced back to early 2000BC when hieroglyphics were used in Egypt to decorate tombs to stylistically tell the life history of the deceased [3]. This area of research was given a name Cryptography, which derives from two (2) Greek words: “Cryptos” and “Graphein” which means hidden or secret and writing respectively. Combining the two,

DOI: 10.14738/tnc.25.504
Publication Date: 16rd October, 2014
URL: http://dx.doi.org/10.14738/tnc.25.504

the meaning now becomes "secret writing or hidden writing". Cryptology could be seen as a practice and study of techniques for secure communication in the presence of third parties [4, 8].

Cryptography in modern days has evolved through many developmental stages, and several cryptographic ciphers have emerged as a result. Each cipher has a distinct algorithm or emerging from the existing one as a higher specification of the existing one. As these ciphers emerge, hackers also undertake research on how to make the efforts worthless, and so on, making this area a research oriented field [4].

Cryptographic algorithms are essential in securing documents on the communication network [14]. The use of multiple algorithms to realize information secrecy enhances the security of the information by requiring several keys before the meaning of information can be revealed in plaintext. Each of the transformations requires a given level, where each level is assigned 1 (one) and the total level for the transformation is given as n, for both encryption and decryption processes respectively. Take for instance, if the total transformation for a given implementation is two, then n is equal to '2'. This paper proposes a generalized algorithm for using more than one transformation cipher on a single plaintext to realize one output. This is termed multilevel cryptography.

2 Related work

Lein Harn and Hung-Yu Lin [5] 1990 proposed a key generation scheme for multilevel data security using bottom-up approach. The term multilevel was used to mean variable securities at different access levels with many users of a single system having different keys at each different access level. This approach was formed modifying the approach proposed by Akl and Taylor [10] 1982 using a top-down model. Usha et al. [13] proposed a multilevel encryption-decryption of text into cipher data in which its characters are encoded uniquely into its corresponding cipher and eliminating the possibility of any pattern as described in their paper titled 'Secure Multilevel cryptography Using Graceful Codes'. It uses more than one level of security by employing many ciphers to disguise any pattern.

Rashmi et al. [9] introduced the culture of securing images using chaotic mapping and elliptic curve cryptography in a network environment. The dependency of stream ciphers on pseudo-stochastic sequences was noted as it can produce a pseudo-random sequence with good randomness. Hardjono and Sebbery [12] discovered a system that makes use of hierarchical keys used to encrypt and decrypt data stored in databases using the RSA cryptosystem with additional restriction of encrypted information to the public. The base of the systems security is discrete logarithms and the term 'multilevel' used in this context means multiple users with different securities.

Multi-Level Crypto Disk: A secondary Storage with Improved Performance was introduced by Chaitanya et al, [11]. They discussed the issue of hard disks becoming increasingly vulnerable to security attacks as they are now accessed remotely, either with mobile devices or in other unanticipated operating environments. They highlighted the demerits of using single data encryption on storage devices, proposing a secure disk using multiple crypto levels.

Multi-Level Cryptographic Functions for the Functionalities of Open Database System was designed and implemented by Adio et al. [1]. This is a secure open database system for an organization that can open their information system for access by different users. The implementation does not require input to be

hidden from anyone or converted to place holder characters for security reasons, but the user only needs to study the sequence of codes and active boxes that describe his password and uses it in place of his active boxes.

A secure information transmission using Multilevel Steganography and Dynamic Cryptography was proposed by Navneet S. Sikarwar [7] in his paper titled 'An Integrated Synchronized Protocol for Secure Information Transmission derived from Multilevel Steganography and Dynamic cryptography'. He juxtaposed the use of both simple steganography and cryptography proposing that multiple and dynamic codes give more security. Maruti et al. [6] presents a practical implementation of a quasigroup based multilevel encryption for data and speech. It makes use of an indexed scrambling transformation for signal authentication, encryption, and broadcasting applications in secrete-key cryptography. The results presented shows that a quasigroup transformation is very effective in destroying the structure of the input signal, and hence can be a good encryption technique.

3 Types of Multilevel Techniques

The major elements of multilevel cryptography are the contributing cryptographic ciphers, which are arranged in a desired sequential order. The term "multilevel" implies that more than one transformation must take place within the system in order to produce an output (cyphertext). The product of a multilevel algorithm is obtained from an organized sequential transformation of input (plaintext) with a desired encryption cipher(s). There are three general classes of algorithms used in multilevel cryptosystems:

a) Same cipher with same key
b) Same cipher with different keys
c) Different ciphers

3.1 Same Cipher with Same Key (SCSK)

When a multilevel cryptosystem is made of the same cipher with the same key, the transformation used can said to be an iteration of a particular cipher. The security of the ciphertext now lies on the complexity of the cipher, key management, and the number of the iterations made. Given that the transformation order number of a multilevel cryptosystem with the same key is **n**, then the reverse computations that can be done with the ciphertext in order to regain the message in clear text is also **n**. As the security of SCSK-multilevel implementation lies on the complexity of the cipher, key management, and the number of iterations made, it is quite certain that the owner of the system need to secure their implementation by hiding the cipher used, the key used and the number of transformation made from the knowledge of adversaries.

3.2 Same Cipher with Different Keys (SCDK)

A SCDK-multilevel structure is said to have been made when the same cipher is used with different keys at different stages. It is similar to SCSK but uses variable keys per iteration. The keys are varied sequentially based on choice, and are kept constant per given implementation. The security of the SCDK-multilevel structure lies on the type of cipher used, number of keys used, key management, and number of iterations made. The sequential order of keys applied per iteration in the SCDK-multilevel structure should be noted, as it has to be reversed during the decryption process.

3.3 Different Ciphers (DC)

When a multilevel technique is enforced with different algorithm, the security of the implementation is high and relies on the complexity of the contributing ciphers, the number of keys used, key management, as well as the number of transformations made. In this case one particular cipher is not used sequentially twice, but can be used after another cipher has been applied, this means that a particular cipher cannot be used for both i-operation and i+1-operation, but can still be used after i+1-operation. This type still has some other subtypes that are determined by the keys used but will not be captured in the general algorithm. The sequential order of encrypting ciphers with their keys should be kept constant as it has to be reversed during the decryption process per every implementation.

4 Methodology

Formal method was adopted to define and formalize the definition of n-cryptographic algorithm. A plaintext was designated as input for the algorithm, while the output is the cyphertext. The transformation of cipher (αi) and key (βi) was done using the initial element i. The formal definition of n-cryptographic cryptosystem is done in the following subsections.

4.1 Order Number of a Multilevel Scheme (n)

The order number of a multilevel implementation (**n**) could be defined as the number of transformations that will take place before producing the desired output (ciphertext). This number of times does not depend on the type of cipher nor upon the key used. For every implementation, '**n**' is placed as the finite-transform-number, while '**i**' is a variable that an increment as the transformation proceeds. For every transformation, the **i**th value increases with **+1,** while it is set to 0 (zero), at the beginning of an operation. The **i**th value defines the termination of the process given that the transformation rules were kept constant.

The termination of the transformation process is said to occur when the **i**th value equals the value of **n.** hence

For the first transformation, $i_1 = 1$,
For the second transformation, $i_2 = i_1 + 1 = 2$,
For the third transformation, $i_3 = i_2 + 1 = 3$,

. . .

For the last transformation, $i_n = i_{n-1} + i_1 = n$.

In a multilevel process, the value of **i** in an uninitiated transform state is **0** (zero), and increments as above.

4.2 N-Cryptographic Algorithm for Multilevel Structure

This concept is made to harmonize the general representation of multilevel cryptographic scheme; the implementation adopted several terminologies in order to describe its structure. Such terminologies are explained below:

i. ***i*** is an incremental variable that determines the present order of operation
ii. ***n*** is the order number of the multilevel structure
iii. α_i is the cipher used per i-operation

iv. α_{i+1} is the next cipher to use after i-operation
v. β_i is the key used per i-operation
vi. β_{i+1} is the next key to use after i-operation
vii. = is the assignment operator
viii. != is the non-equal-to operator

The n-cryptographic algorithm for multilevel techniques is shown below, and the algorithm is presented in Figure 1:

Table 3.1 Steps in Generalized Multilevel Scheme

1	Begin	9	$\beta_i = \beta_{i+1}$; $\alpha_i = \alpha_i + 1$
2	Take Plaintext	10	Else if (β_i != β_{i+1})
3	Perform i-transform with α_i-Cipher and β_i-Key	11	$\beta_i = \beta_{i+1}$
4	Do	12	Else
5	Display Ciphertext	13	i = i + 1
6	While i > 1 AND i == n	14	Move to Line 3
7	Else	15	End
8	If (α_i != α_{i+1})		

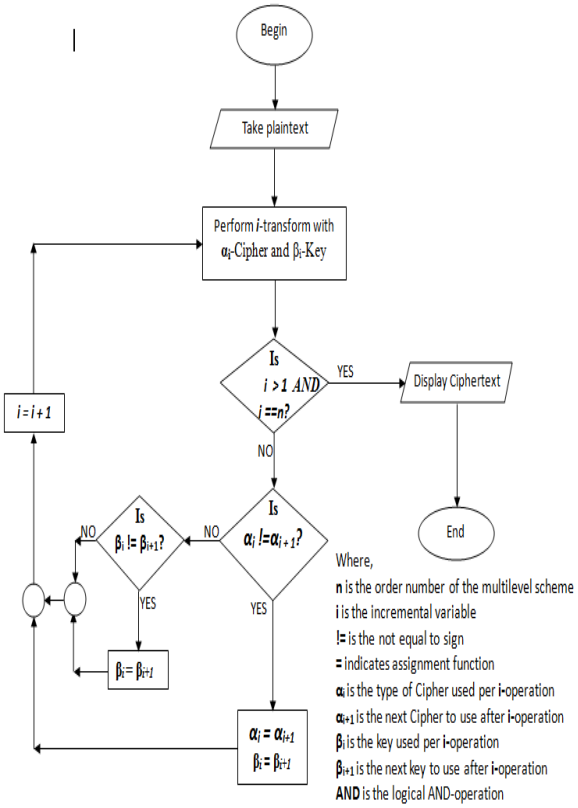


Figure 1: Flowchart Representation of the Generalized Multilevel Scheme

4.3 Analyzing the security capability of Multilevel Cryptography

As stated earlier, the security of a multilevel cryptosystem is dependent on the complexity and the structure of its component cipher(s) which can be made more stringent by using more than one type of

cryptosystem or the same cipher with different keys. It also depends on the key-length: the length of a given key has much to do with the security of a multilevel structure. It affects the possibility of factoring the key components as well as the possible permutations that can take place in order to realize the plaintext. The order-number of the multilevel cryptosystem, **n**, determines the number of transformations that was performed per given implementation, and helps to keep the implementation tight. The security of a multilevel structure also depends on the complexity and type of the cipher. This determines whether the public has the key of a particular individual or if it is only a selected partner as may be the case for asymmetric and symmetric key cryptosystems respectively.

Key management could also be seen as a serious security implication in a multilevel state. For instance, in an SCSK implementation with symmetric cipher, key revelation could be a great challenge as continuous trials of reverse computation could reveal the ciphertext in clear text. The type of programming language used for the implementation of the multilevel structure also has a great impact on the security of the system.

5 Cryptanalysis: Possibility of Multilevel Encryption

In a multilevel state, there still exist possibilities of cryptanalyzing the ciphertext (output). These possibilities could be caused mainly by the factors below:

a. Weakness of the selected cipher(s)
b. Weakness of the programming language used for the system construction.
c. Key management
d. Number of transformations made (order number of the multilevel scheme)
e. Number of cipher(s) used

5.1 Weakness of the selected algorithms

In cryptography, most of the often known algorithms have their corresponding weaknesses. These weaknesses have made it variably possible to break the security it ensures and hence keeps the study (cryptography) dynamic. The weakness associated with a particular cipher is different from the one associated to another, and depends on the type of cipher. For instance, if an implementation contains shift-cipher under the modulus of 26; the possible permutations that can be made to realize the security at that level is 26, which is very small to ensuring that the security of information is kept.

5.2 Weakness of the Programming Language used for the System Construction

The type of programming language selected for the implementation helps to ensure the total security of the encryption scheme in an application level environment. This is dependent on the security of the construct made with the programming language. Some programming languages are prone to attacks such as: buffer overflow attack, SQL injection attack etc.

These attacks can also be realized if it is used for the implementation of multilevel scheme, so it is advisable to use a programming language that is devoid of attacks for the implementation of multilevel cryptography.

5.3 Key Management

Key management is paramount in ensuring the security of multilevel scheme; this has to be built upon trust of the individuals involved on the communication. Some cryptographic ciphersare not public key based and hence are not supposed to be revealed to the public except to those involved in the communication. Thus multilevel cryptographic keys should be kept secret among those that use the implementation.

5.4 Number of Transformations Made (Order Number of the Multilevel Scheme)

The order number of a multilevel implementation determines the number of encryption operations that have taken place or that will take place per that particular implementation. This also shows how many reverse computations with the ciphers that will take place before the ciphertext can be realized in clear text. Making this number higher helps to achieve a very high level of security. In fact, one of the major features that helps to make a multilevel scheme different from other methods is the ability to keep this order number high with a single implementation. As you could see from the flowchart figure above (Figure 1), the multilevel scheme could only be satisfied if the i-variable is greater than one. Thus the i-operation can only satisfy this condition of becoming equal to n when the value is two and above. Keeping this n value secret also determines the security of a given implementation.

5.5 Number of Cipher(s) Used

The number of ciphers used per given implementation is another factor that influences the security of the scheme. If an implementation contains a single type of cryptosystem, when the weakness of the particular cipher is broken, the entire system is broken; but if it contains more than one type of cipher, breaching of one component cipher does not break the system entirely. So using more than one particular cipher is preferable, especially using an implementation that involves both private and public ciphers. The beauty of involving both private and public ciphers in one implementation cannot be over emphasized as it helps to make a multilevel system more resilient against cryptanalysis attack.

6 Future Research

The future direction of this research is to implement this n-cryptographic multilevel algorithm and examine its effectiveness and efficiency against the existing methods.

7 Conclusion

This research has shown how two or more encryption schemes can be combined to be more effective. The research proposed and examined the benefits and weaknesses of the blended algorithm for multilevel encryption. The success of information security lies on the inability of adversaries to understand the message if intercepted on a communication network. Several contributions have been made in this regard; formulating cryptographic ciphers that helps to transform the plain information into unintelligible format, as this has not realized perfect information secrecy. As the realization of perfect information secrecy remains a dream, using the proposed multiple cryptographic ciphers to transform given information helps to increase the difficulty of cryptanalyzing encrypted information into its plaintext. The proposed algorithm does not give preference to any particular cipher; it presents an avenue for the possibility of such implementation and also classifies the possible implementation according to types. Multilevel cryptography implements multiple cryptographic ciphers onto a single plain text. The plaintext is taken as shown in the flow diagram and transformed with the predetermined

ciphers until the order-number of the implementation is reached. The higher the order number of a given implementation, the more secure the resulting ciphertext will be.

ACKNOWLEDGMENT

The researchers wish to acknowledge Dr Andrew Fluck of the Faculty of Education, University of Tasmania, Locked Bag 1307, Launceston, TAS 7250, Australia for editing this work.

REFERENCES

[1]. T. A. Akinwale, F. A. Adekoya, and E. O. Ooju, "Multi-Level Cryptographic Functions for the Functionalities of Open Database System". *Computer Technology and Application 2 (2011), Pp. 730-735*, 2011.

[2]. J.S. Gustavus, "Symmetric and Asymmetric Encryption". *Computing Survey.* Vol. 2 (4), pp. 321, 1979.

[3]. J. N. Ugwu, "Multilevel Offline Cryptography Support System". Undergraduate Project, Federal University of Technology, Minna, Nigeria, 2014.

[4]. W. Judy "Notes from her Math 398 course taught in the Spring of 2002 at UNL". Ericsson AB, ERLANG Secure Socket Layer 5.1.1., 2002.

[5]. L. Harn and H. Lin, "A cryptographic Key Generation scheme for multilevel Data Security". *Computers & Security, 9 (6) 539-546***.** Computer Science Telecommunication program, University of Missouri-Kansas City, Kansas city MO, U.S.A, 1990.

[6]. M. Satti and K. Subhash, "Multilevel Indexed Quasigroup Encryption for Data and Speech". *IEEE Transaction Broadcasting 2009*. Authorized Licence use to: Oklahoma State University, 2009.

[7]. N. S. Sikarwar, "An Integrated Synchronized Protocol for Secure Information Transmission derived from Multilevel Steganography and Dynamic Cryptography". *International Journal of Computer Science and Telecommunication, Vol. 3(4)*, 2012.

[8]. A.M. Richard "Codes The Guide to Secrecy from Ancient to Modern Times, Discrete Mathematics and Its Applications'. Taylor & Francis Group, LLC Chapman & Hall/CRC, an imprint of Taylor & Francis Group, 2005.

[9]. R. K. Gawande, P. S. Kulkani and K. A. Ganar, "Multilevel Image Encryption using Chaotic Mapping and Elliptic Curve Cryptography". *International Conference of engineering Innovation and Technology.* ISBN: 978-93-81693-77-3; Nagpur, 2012.

[10]. S. G. Akl and P. D. Taylor, "Cryptographic solution to a multilevel security problem". Proc. Cryp~o-82, Sanla Barbara, CA, AU~USI 23-25, pp. 237-250, 1982.

[11]. S. Chaitanya, B. Urgaonkar, A. Sivasubramaniam, "Multi-Level Crypto Disk: Secondary Storage with Improved Performance vs Security Trade-offs". *Technical Report CSE-09-006,* 2006.

[12]. T. Hardjono and J. Seberry, "A multilevel Encryption Scheme for Database Security", Department of Computer Science, University College, The University of South Whales. Australian Defense Force Academy Canberra, A.C.T., 2600, 1989.

[13]. D. G. Usha and R. S. D. Wahida Banu, 'Secure Multilevel Cryptography Using Graceful Codes'. *International Journal of Information and Electronics Engineering, Vol. 2(5).* 2012.

[14]. O. S. Adebayo, V. O. Waziri (PhD), J.A Ojeniyi, S. A. Bashir, A. Mishra, 'Information Security on The Communication Network In Nigeria Based On Digital Signature'. *International Journal of Computer Science & Information Security (IJCSIS)'.* Vol. 10 (11), pp. 57-63. ISSN 1947-5500, 2012.

TNC Transactions on Networks and Communications
SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
VOLUME 2, ISSUE 5
ISSN: 2054 - 7420

Central Locker System for shopping mall using NFC Based Smartphone

Siddarth Poddar
System LSI Group, Samsung R&D Institute (SRI-B), Bangalore, India
s.poddar@samsung.com

ABSTRACT

Depositing purchased item at each store's locker system and carrying a token in hand for the assigned locker while entering a store in a shopping mall as well as collecting the item back while exiting the store is a time consuming exercise and even long delayed if the token is lost. The time would even multiply visiting multiple numbers of stores in a mall. Sometimes waiting in a long queue for depositing item at the store is an irritating hassle and everyone wants to skip this hassle. This paper introduces a novel solution using NFC enabled Smartphone and NFC reader located at each checkout counter of the store which provides an easy and convenient way to immediately keep your purchased things safe at the centrally located locker system and move around openly. Locker number would be generated at the store checkout counter after purchasing the item and just a tap of smartphone on NFC reader save the locker number information on NFC application installed on smartphone, which is used to retrieve the item back while exiting the mall. With the proposed solution, usage of extra space for each store's locker system and the corresponding staff can be eliminated, leading to cost reduction, promote efficiency and enhance customer service experience.

Keywords—Near Field Communication, Host based card emulation, Shopping mall, Central locker system.

1 Introduction

Shopping is an activity which allows one or more retailers to present their goods and products to the customers to help them in selecting a suitable item according to their desire [1]. Shopping mall is providing a bigger platform, to have all the retailers, merchandisers at one place selling their products, and to the customers providing wide variety of products all at one place with different brands. It can be a very pleasant or awful shopping experience for customers depending on number of factors like their convenience, easy access to stores/outlets etc. These factors would be even more significant when it comes to those who are over particular or have shopping addiction. Customers have to carry purchased items in hand while moving from one store to another or even roaming inside the mall. Customer need to deposit the item at the counter of each store while moving in. Store guy provides a token for that to collect it back while exiting the store. This consumes a lot of time in order to deposit the item and to get it back. Some shopping malls have a system of central locker system, in which customer has been provided with a token at the checkout counter of the store only while purchasing the item which can be used to retrieve the item back while exiting the mall . It allows them to move openly and even saves a

DOI: 10.14738/tnc.25.520
Publication Date: 25rd October, 2014
URL: http://dx.doi.org/10.14738/tnc.25.520

lot of time. Both such processes have at least one limitation; customer should not miss the token. To overcome these limitations would be greatly valuable for shopping mall business sector to provide better service experience to customer, helps in cost reduction, and better space utilization. This paper presents a novel solution using NFC technology [2]. NFC has already shown its capabilities in many areas like communication, social networking, entertainment and even in the field of shopping business like mobile payment, indoor navigation system, customer assistance etc. [3, 4] In addition to all these, one more application of NFC in shopping mall is the locker management system which helps the customer to move around openly without carrying any item in hand and not even the token inside the mall. This comprises of centrally located locker system at the entrance or at the exit door of the shopping mall, NFC enabled smartphone with customer, NFC reader [5] at the checkout counter of each store of the mall as well as at the Parcel Dispatch Station. The proposed solution starts with a tap of customer's NFC enabled smartphone on the NFC reader placed at the store checkout counter which stores the assigned locker number of his purchased items to the NFC application in the phone. The saved locker number is used to retrieve the item back from central locker system while exiting the mall.

2 Near Field Communication (NFC)

NFC technology is a bidirectional, contactless communication between two devices. It operates at RF frequency of 13.56 MHz in a short range of 10cm or even less with a data rate ranging from 106kbit/s-424kbit/s [2]. When the two NFC devices come in close proximity triggers the communication, and NFC is just a platform for communication between two devices. As the technology operates in a very limited range, it is ideal for secure transactions [6]; even it serves as a safeguard against hackers. NFC can operate in three communication modes:

a) Tag Read/Write mode;
b) Peer to Peer mode;
c) Card Emulation mode.

Card emulation mode enables NFC based smartphones to act as a contactless card or tag to interact with NFC reader. The card or tag application lives on some other piece of hardware like Embedded Secure Element (ESE), UICC etc. that is wired to the NFC controller [7]. A NFC based smartphone does not generate its own RF field; instead NFC reader generates this field. Supported RF interfaces are ISO/IEC 14443 Type A, ISO/IEC 14443 Type B and Felica. NFC reader can write any data to a tag/NFC device or read from it. Data can be any simple text message, or URL etc. NFC forum has suggested a common specification for all these type of data called NDEF format (NFC Data Exchange Format). NDEF specifications defines a message encapsulation format to exchange information, e.g. between an NFC Forum device and another NFC forum device or an NFC Forum tag [8]. Android 4.4 introduces an additional method of card emulation that does not involve a secure element, called hostbased card emulation (HCE). This allows any Android application to emulate a card and talk directly to the NFC reader. In this case, the data is routed to the host CPU on which Android applications are running directly, instead of routing the NFC protocol frames to a secure element.[9] The proposed solution operates in NFC host based card emulation mode using NFC based smartphone as a NFC tag to store the information and the NFC reader. NFC readers are to be placed at the store checkout counter as well as the parcel dispatch station. NFC reader will generate RF field and start sensing NFC capable devices in its vicinity. When mobile device is tapped on NFC reader, it senses the device and starts with the initialization and anti-collision process, and automatically activates the specific installed application on

the device and store the data into it which would be visible to customer on screen to verify. At the time of exiting mall, customer is advised to tap the device again on NFC reader placed at the parcel dispatch station, which again automatically activates the specific installed application on the device. At this time, NFC reader will read the data stored in the application instead of writing. The valid data stored would be visible to customer on screen for verification. Customer needs to verify the items and Exit session on device. Even, Customer can check the locker numbers and other details anytime while opening the application installed on device.

3 System Design

A system design has been proposed in this paper for the customers having NFC enabled smartphone to store the information of the locker which will have purchased product of the customer. This Application is intended for deployment in Shopping Mall to be used by customers during shopping. It involves both the customer as well as the seller. This system design has only one assumption that the wireless connection (Wi-Fi) inside the shopping mall should be reliable and fast. It comprises of three components:

a) Customer's NFC enabled smartphone
b) Locker Generate system at each checkout counters of stores.
c) Parcel dispatch station at the central locker system of mall where the purchased items are to be kept.

a) NFC Enabled Smartphone:

The most important component of this solution is NFC enabled smartphone with customer. A couple of years ago there were no more than a handful of phones supporting NFC. Now, 9 out of 10 of the top ten manufacturers sell NFC phones - the exception being Apple [10]. Mobile device should have the specific application intended for this solution installed. If application is not installed already, then mobile device should be configured to redirect to the application web for downloading using Wi-Fi. The solution prototype is developed on the Android smartphone in eclipse using Java along with the libraries provided by the Android SDK [11]. Customer need to register first to get the username or user id which helps to maintain his data in the shopping mall database. Even it helps customers to check the details of assigned lockers anytime as well as the previous history also. The information stored in the application comprise of Store name, Assigned Locker, number of Packets, cash memo id and status. The information is stored in the two different structure in the mobile application based on the "status" information. If the status of entry is "In Store" which means the packets are there in the central locker system of the shopping mall will be stored in application under the button named as "My Locker". It can be used for convenient tracking of purchases. It is completely implementation specific and can be modified as per the need.

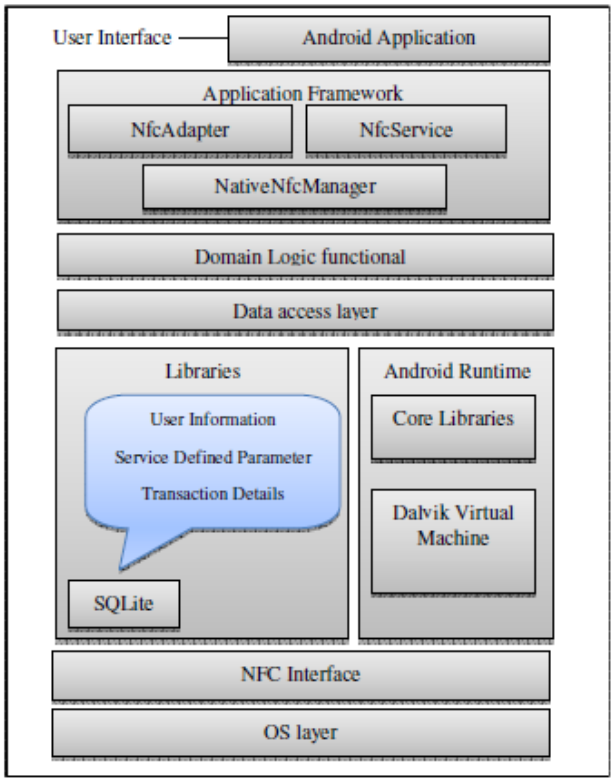


Figure1. NFC Smartphone Application Software Architecture

b) Locker Generate System:

Locker Generate System is located at each store which comprises of NFC reader connected to the seller's system via USB interface. When asked to generate a locker, seller needs to enter number of packets (purchased by customer) and send this information to the shopping mall sever (database) which will, in response, generate locker number(s) depending upon the number the packets. Lockers can be of different types based on the requirement by a particular store. Shopping mall database has the complete information of the lockers available at the locker system. Cash memo id is assigned to the assigned lockers in the database of shopping mall and can be referred in case of any error while collecting the packets. After generating the locker number(s), user is asked to tap smartphone to the NFC reader placed on the counter. NFC reader writes the data (locker number) to the NFC tag built in the smartphone. If the data has been written successfully, popup message will come on screen showing the detailed information. Database will maintain a status of each entry with status as "In Store" which mean the packets are there in the central locker system of shopping mall.

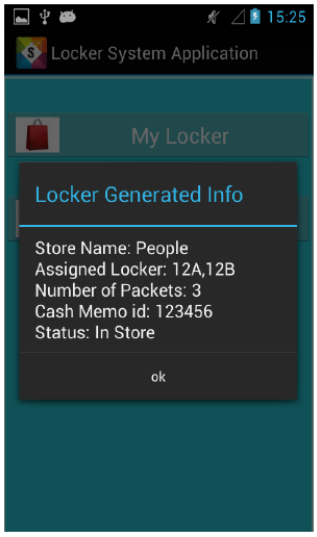


Figure 2. Information popup and stored in smartphone application database at the time of locker generation at store checkout counter

c) Parcel Dispatch Station:

Parcel dispatch station is located in the central locker system at the entrance or exit of mall which is used when the customer would like to get back his/her purchased items stored in the locker system. The dispatch station will have NFC reader connected to the system via USB interface. The NFC reader is used to read the information from the NFC tag built in the smartphone. When customer taps on it, the NFC reader only read the data whose status is "In Store" which is a valid data for shopping mall database and the data will be send to the shopping mall database which will in turn displays the complete information on screen. Status now of each entry change to "Delivered" once it is verified successfully by customer. While tapping, customer gets all the details on the UI showing "Verify and Exit Session" button. On pressing the button, the database in mobile is also get updated.

Figure 3. Information popup and stored in smartphone application database at the time of collecting packet from Parcel Dispatch Station

4 Service Outline

The service outline for this proposed solution is described in five steps as follows:

***Step 1**:* At the time of billing an item from a store, seller will ask the customer to avail the facility of locker system. If customer agrees to use the facility, Seller will enter the information (like the number of packets of the purchased item) and send it to the shopping mall server (database) which has the information about the available lockers at the parcel dispatch station and will generate a locker number depending on the packet classification and the number of packets. Corresponding to the generated locker number, details of number of packets, cash memo id and status will be stored in the shopping mall database. The status information would be set as "In Store", which means the packets are stored in parcel dispatch station.

***Step 2**:* Customer will be asked to tap the NFC enabled smartphone to the reader placed at the checkout counter of the store. It will automatically activate the locker system application, which is specifically developed for this service, if it is pre-installed on the smartphone or else it will show the direction to the customer to download and install the application using Wi-Fi.

***Step 3**:* While tapping on the reader, the reader writes the details of locker number on the application in the phone as well as displays it on the screen.

***Step 4**:* After locker number gets stored in the phone and the details are verified by the customer, seller will deliver the packets to the parcel dispatch station and customer can move ahead to other stores for further shopping.

Step 5: When the customer approaches the parcel dispatch station while exiting the mall, customer will be directed to touch the smartphone on the reader placed at the counter; reader will then read the captured data from the application and send it to the database for verification. Database in return sends the information about the valid entries and displays it on the screen. The displayed data will have cash memo id and all the locker numbers assigned to the customer from all the stores of the shopping mall for the purchased items. Now this time the status of the items is updated from "In Store" to "Delivered". Customer can collect the packets and verify the items. In case of any mismatch, cash memo id can be used as a reference to track the item.

***Special Case**:* There can be a situation when customer mobile device may be temporarily un-operational. For e.g., the battery may be drained or the mobile application is not responding or even the NFC reader at the parcel dispatch station has some issue. In such cases customer can use the cash receipts of purchased items to retrieve them back from parcel dispatch station. After collecting the packets, the "Status" information of the entries in the shopping mall database is updated from "In Store" to "Delivered". But the status of the corresponding entries in the customer's mobile application database will be unchanged as "In Store". So, the next time customer uses his mobile phone to avail locker facility and use the mobile device to retrieve packets from parcel dispatch station, the database of mobile device application is updated and bring in sync with the shopping mall database. Alternatively, Customer can even update his mobile database later on while connecting it to the server.

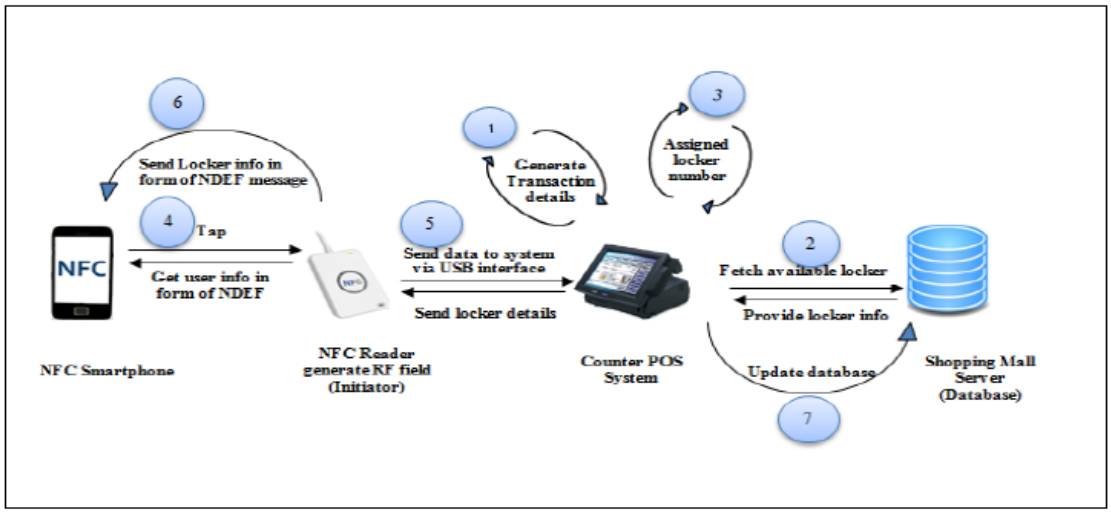


Figure 4. Solution Architecture- Locker Generation at store checkout counter

There are several benefits to customers of using this central locker system solution in shopping mall:

- Security of their purchased products.
- Organization (no more token lost!).
- Record Keeping.
- Minimize time consumed for depositing packets at store locker system while entering or exiting store.
- Eliminate chances of even losing purchased item inside mall.
- Get rid of the overhead of carrying packet in hand while roaming inside the mall.

As for the shopping mall retailers, it offers few useful benefits:

- NFC reader can be easily integrated with any POS (point of sale) system.
- Reduce Locker system installation at each store.
- Reduce Man-power also.
- Provide better service quality to customer.
- Allow them to convert their casual customers to loyal customers.

5 Conclusion

Customers like to visit stores that offer something different. If it is something that makes life more convenient for them, then they are more likely to revisit such a store [12]. This is exactly NFC technology offers in this paper. This paper presents an easy and convenient solution to customers for not carrying any purchased items in hand and storing it safely, while inside the mall. It has a very minimal customer interaction only, i.e., just a tap on the NFC reader. It will help to keep the track of items purchased from the shopping mall. It involves a lot of work for the shopping mall management at the backend to maintain the database and to have a large locker system area with the locker system of each and every store maintained separately at central locker system to easily retrieve the stored item back. It also poses a challenge to deliver packet from store to parcel dispatch station. The design of central locker system

as well as delivering of packet to dispatch station is completely shopping mall specific. For customer, it's just to download the application in their smartphone and use it while stepping inside the mall. This would even show many advantages to business sector also in terms of cost reduction, reducing man power, better space utilization, i.e., one central locker system instead of having locker system as each and every counter. This solution presents a win-win situation for both retailers and customers as far as benefits are concerned. For future work, this solution can be further used to support marketing and promotions. It can be used to integrate coupons, offers and loyalty programs which can be sent to customer's smartphone while using this service based on their past purchases. It does not require any other customer signup or registration or any overhead. With the emergence of NFC technology in Smartphones and its applications in the service sector business; this solution can further improve the usability of these systems and provide flexibility to the shopping mall management system to improve service experiences using the solution framework.

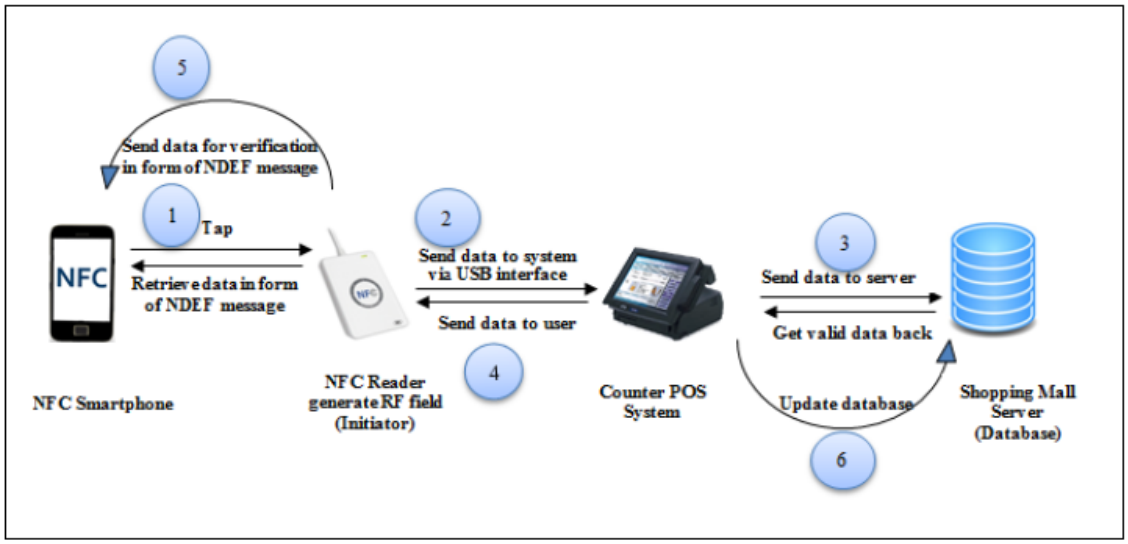


Figure 5. Solution Architecture- Parcel Dispatch Station

ACKNOWLEDGMENT

I would like to extend my sincere thanks to Mr. Pritam Bhowal, Lead Engineer, Samsung Research Institute-Bangalore (SRIB) for his kind co-operation in analyzing the key features regarding the project and provide necessary information. Many thanks and appreciations to him for helping me in completion of this paper.

REFERENCES

[1]. Shopping, http://en.wikipedia.org/wiki/Shopping [referenced 20th May, 2014]

[2]. Near Field Communication, http://en.wikipedia.org/wiki/Near_field_communication [referenced 25th May, 2014]

[3]. Indoor Navigation System, http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5954491&url=http%3A%2F%2Fieeexplore.ieee.org%2Fstamp%2Fstamp.jsp%3Ftp%3D%26arnumber%3D5954491

[4]. Mobile Payment, Components for an interoperable NFC mobile payment ecosystem http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=6478871 [referenced 17th July,2014]

[5]. NFC Reader, http://www.acs.com.hk/en/products/3/acr122uusb-nfc-reader/ [referenced 18thMay 2014]

[6]. http://businesstoday.intoday.in/story/near-fieldcommunications-nfc/1/21671.html [referenced 18th June 2014]

[7]. http://randomoracle.wordpress.com/2013/12/02/nfc-cardemulation-and-android-4-4-part-i/ [referenced 17th July, 2014]

[8]. NDEF technical Specification, http://www.eetchina.com/ARTICLES/2006AUG/PDF/NFCForum-TSNDEF.pdf?SOURCES=DOWNLOAD [referenced 18th June 2014]

[9]. https://developer.android.com/guide/topics/connectivity/nfc/hce.html [referenced 17th July, 2014]

[10]. http://rapidnfc.com/nfc_enabled_phones [referenced 18th June 2014]

[11]. Android Developers, http://developer.android.com/index.html.

[12]. http://www.geekbusiness.com/2012/10/let-your-retail-businesstake-advantage-of-nfc-technology

TNC Transactions on Networks and Communications
SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
VOLUME 2, ISSUE 5
ISSN: 2054 - 7420

Fractal Antennas
A Novel Miniaturization Technique for wireless networks

Abdelati REHA[1], Abdelkebir EL AMRI[2], Othmane BENHMAMMOUCH[3] and Ahmed OULAD SAID[4]
[1,2,3]RITM Laboratory, ESTC, Hassan II University, CASABLANCA, MOROCCO;
[4]Royal Air Academy, MARRAKECH, MOROCCO
[1]reha.abdelati@gmail.com, [2]elamri_abdelkebir@yahoo.fr, [3]othmane.benhmammouch@gmail.com,
[4]a_ouladsaid@hotmail.com

ABSTRACT

In the recent years and with the multiplication and miniaturization of telecommunications systems and their integration in restricted environments, such as Smart-phones, tablets, cars, airplanes, and other embedded systems. The design of compact multi-bands and Ultra Wide Band (UWB) antennas becomes a necessity. One of the interesting techniques to provide this kind of antenna is the use of fractal structures.

Keywords: Fractal antennas, Broad-Band antennas, Ultra Wide Band Antennas, Multi-Band antennas, wireless communications.

1 Introduction

The numerous applications of telecommunication to the advances of technology have necessitated the exploration and utilization of most of the electromagnetic spectrum. Also, the advents of broadband systems have demanded the design of broadband and multi-band antennas. In addition, the use of simple, small, lightweight, and economical antennas, designed to operate over the entire frequency band of a given system, would be most desirable. In recent years, one of the techniques used to design this kind of antenna is the use of the fractal structures.

The term “Fractal” means linguistically “broken” or “fractured” from the Latin “fractus.” This term was created by Benoît MANDELBROT 40 years ago in 1974.

Fractals are geometric shapes, which cannot be defined using Euclidean geometry, are self-similar and repeating themselves on different scales like clouds, mountains, coasts, lightning, etc. [1, 2].

The fractal geometry has been applied to many fields such as:

- Medicine: structure of the lungs, intestines, heartbeat,
- Meteorology: clouds, vortex, ice, rogue waves, turbulence, lightning structure,
- Volcanology: prediction of volcanic eruptions, earthquakes.
- Astronomy: the description of the structures of the universe, craters on the Moon, distribution galaxies. [3].

DOI: 10.14738/tnc.25.566
Publication Date: 25rd October, 2014
URL: http://dx.doi.org/10.14738/tnc.25.566

Also, fractal geometry has been used in the electromagnetic, and especially in the design of antennas. Several studies have adopted fractal structures and showed that this technique can improve the performances of the antenna and it is one of the techniques to design antennas with multi-band and broad-band behavior [4].

In this paper, we give some generalities of fractal geometries and their dimensions, after that, we describe some linear fractal geometries such as KOCH, SIERPINSKI, DRAGON, TREE, CIRCULAR, CANTOR SET, HILBERT, MINKOWSKI, and finally we discuss the applications of these geometries and their performances in the design of the miniaturized antennas.

2 The Fractal Dimension

Usual dimensions used are integer values. For example, the dimension of the line is 1; the dimension of a cube is 3. For fractal geometries, the dimension used is not necessarily integer value, but we use HAUSDORFF dimension [5]. A fractal consists of smaller replicas of itself. Its HAUSDORFF dimension can be calculated as follows:

$$d = \frac{\ln(n)}{\ln(h)} \tag{1}$$

With: the fractal consists of (n) copies whose size has been reduced by a factor of h.

Here is an example of calculating a HAUSDORFF dimension of KOCH Fractal.

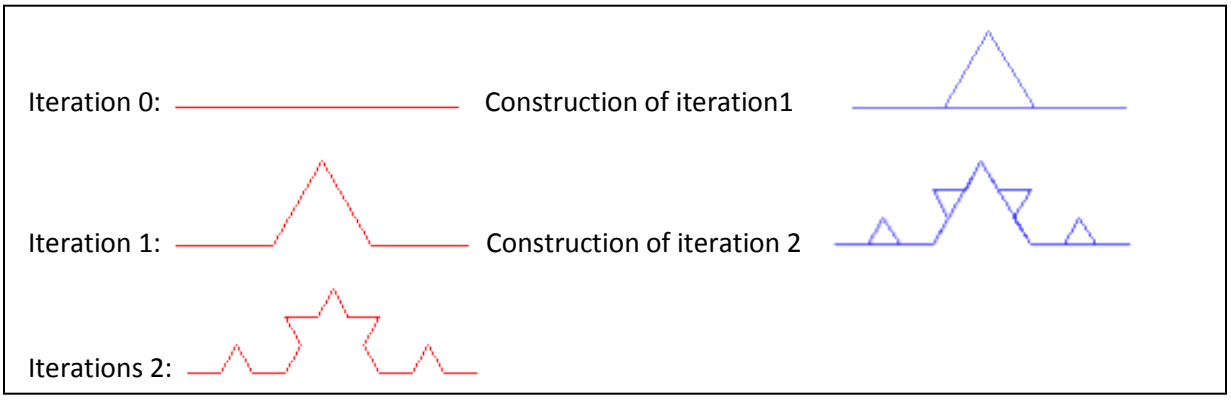


Figure 1: Construction of KOCH fractal iterations

As shown in the figure 1, the geometry of the first iteration is made by four copies of the basic geometry (iteration 0), so n =4. Also, the lengths of the segments making up the geometry of the first iteration, are reduced by a factor of 3, so h = 3.

The HAUSDORFF dimension of KOCH Fractal is:

$$d = \frac{\ln(n)}{\ln(h)} = \frac{\ln(4)}{\ln(3)} = 1.26 \tag{2}$$

3 Types of Fractals

Fractals are classified among three major categories.

- Linear: based on the iteration of linear equations (HILBERT, KOCH, SIERPINSKI, Dragon ...),
- Nonlinear: based on the iteration of complex numbers (MANDELBROT, JULIA ...),
- Random: based on the introduction of a random parameter in the iteration to obtain irregular shapes (such as mountains or clouds)

4 Linear Fractal Geometries

In this section we study the famous fractal geometries, the methods used for the generation of different iterations and calculating their Hausdorff dimensions.

4.1 The KOCH Structure

This structure was invented by the Swedish mathematician HELGE VON KOCH in 1906 before the invention of the term "fractal". There are several variations of this structure:

4.1.1 The KOCH Curve

As shown in Figure 2, the construction of this curve is made from a segment by applying the following steps:

1. The segment is divided into three segments of equal length.
2. An equilateral triangle whose base is the middle segment of the first stage is constructed.
3. Segment which was the base of the triangle of the second step is eliminated.

After these three steps, the resulting object has a shape similar to a cross section of a witch hat.

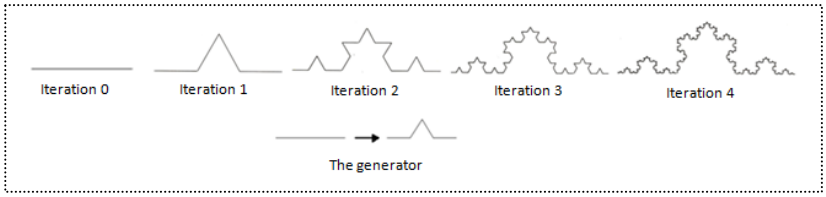


Figure 2: the 4 iterations of the KOCH Curve

4.1.2 The KOCH snowflake

The procedure to build the Koch snowflake is the same as that used for the construction of the KOCH curve except that the base is a triangle, which means that the procedure is repeated three times for each iteration (Figure 3).

The HAUSDORFF dimension of the KOCH structure is given by the equation (2).

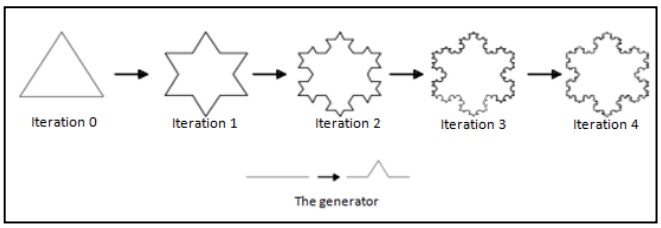


Figure 3: the 4 iterations of the KOCH Snowflake

4.2 The SIERPINSKI Structure

This structure was invented by Polish mathematician SIERPINSKI. There are several variations of this structure:

4.2.1 The SIERPINSKI triangle

The construction of this triangle is made from a solid equilateral triangle and applying the following steps:

a) An equilateral triangle is built and will be taken as a base.
b) Subdivide it into four smaller congruent equilateral triangles and remove the central one.
c) Repeat step 2 with each of the remaining smaller triangles.

The Figure 4 shows the first four iterations of the SIERPINSKI triangle.

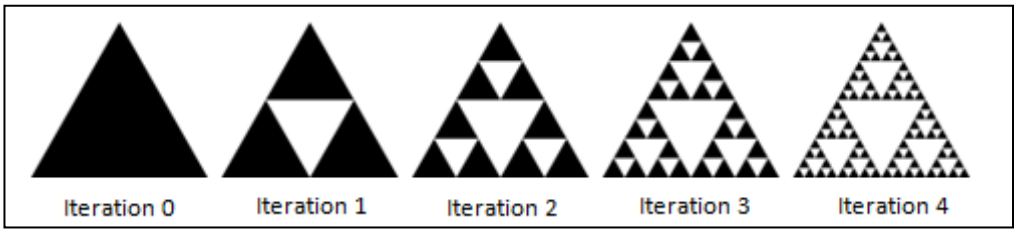


Figure 4: the 4 iterations of the SIERPINSKI Triangle

The HAUSDORFF dimension of the SIERPINSKI Triangle is given by the equation (3).

$$d = \frac{\ln(n)}{\ln(h)} = \frac{\ln(3)}{\ln(2)} = 1.58 \quad (3)$$

4.2.2 The SIERPINSKI carpet

The construction of this structure is made from a solid square and applying the following steps:

a) The square is cut into 9 congruent sub-squares in a 3-by-3 grid
b) The central sub-square is removed.
c) The same procedure (1 and 2) is then applied recursively to the remaining 8 sub-squares.

The Figure 5 shows the first four iterations of the SIERPINSKI Carpet.

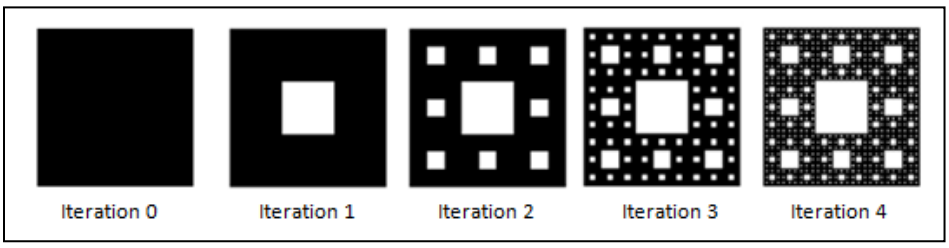


Figure 5 : the 4 iterations of the SIERPINSKI Carpet

The HAUSDORFF dimension of the SIERPINSKI Carpet is given by the equation (4).

$$d = \frac{\ln(n)}{\ln(h)} = \frac{\ln(8)}{\ln(3)} = 1.89 \quad (4)$$

4.3 The Dragon structure

The dragon's name comes from the fact that, for high iterations, the shape of the structure is close to that of the Dragon. The construction of this structure is made from a simple line and applying the following steps:

a) For the first iteration (Figure 6)

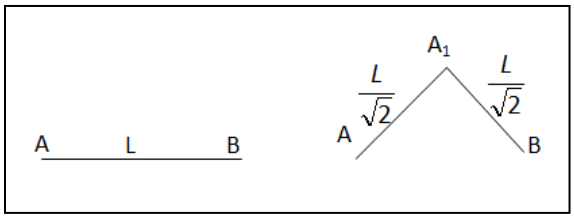


Figure 6: Generation of the first iteration of the Dragon structure

- We move from the segment [AB] to the segment [AA_1] by a rotation with the centre A and the angle $\frac{\pi}{4}$ followed by a scaling whose center is A and ratio $\frac{\sqrt{2}}{2}$;
- We move from the segment [AB] to the segment [A_1B] by a rotation with the center B and the angle $-\frac{\pi}{4}$ followed by a scaling whose center is B and ratio $\frac{\sqrt{2}}{2}$;

b) For the other iterations, we apply the same procedure on each segment. The figure 7 shows the iterations and the 20th iteration of the DRAGON structure.

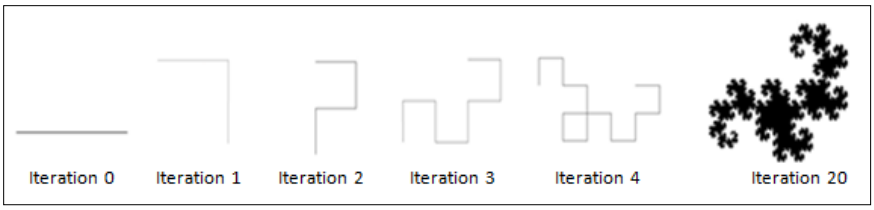


Figure 7: the four first iterations and the 20th iteration of the DRAGON structure

The HAUSDORFF dimension of the DRAGON structure is given by the equation (5).

$$d = \frac{\ln(n)}{\ln(h)} = \frac{\ln(2)}{\ln(\sqrt{2})} = 2 \quad (5)$$

4.4 The Tree Structure

This structure has the same shape of a Tree; there are several kinds of this structure:

4.4.1 The Tree

To generate this kind of fractal structure, we apply the following steps:

1. For the initiator or "iteration 0", the structure has 3 branches, the vertical one is the "parent" the 2 others are the "Childs" with an inclination (θ).
2. At each iteration, the same shape is generated with a reduction factor "h" (Figure8).

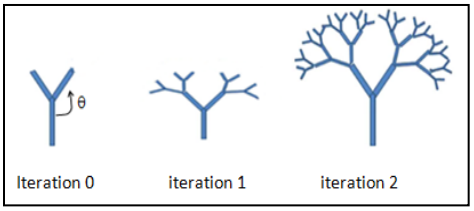


Figure 8: The three first iterations of the TREE Fractal structure

The HAUSDORFF dimension of the Tree structure is given by the equation (6).

$$d = \frac{\ln(n)}{\ln(h)} = \frac{\ln(4)}{\ln(h)} \tag{6}$$

If the reduction factor h=2, the HAUSDORFF dimension is given by the equation (7).

$$d = \frac{\ln(n)}{\ln(h)} = \frac{\ln(4)}{\ln(2)} = 2 \tag{7}$$

4.4.2 The H-Tree

The H-TREE geometry is a modified Tree geometry with the same concept (Figure9). The initiator is a structure like the letter "H". On each iteration, we create 4 copies of the previous iteration with a reduction factor "R".

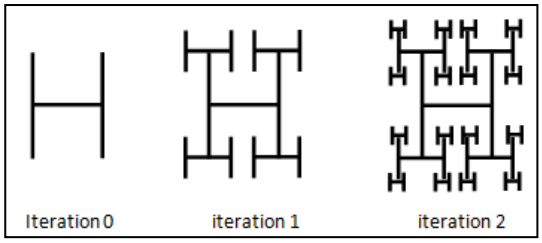


Figure 9: the three first iterations of the H-TREE Fractal structure

For the H-TREE structure, the HAUSDORFF dimension is given by the equation (8).

$$\mathrm{d} = \frac{\ln(\mathrm{n})}{\ln(\mathrm{R})} \tag{8}$$

If the reduction factor R =2, the HAUSDORFF dimension is given by the equation (9).

$$\mathrm{d} = \frac{\ln(\mathrm{n})}{\ln(\mathrm{R})} = \frac{\ln(4)}{\ln(2)} = 2 \tag{9}$$

4.4.3 PYTHAGORE Tree

This structure is constructed with squares. It is named "PYTHAGORE" because each triple square in touch creates a right triangle. To construct this fractal structure, we apply the following steps:

1. we built a simple square,
2. On this square, we construct two other squares, each one is smaller by a factor of $\frac{\sqrt{2}}{2}$, such as the corners of the squares are in contact.
3. The procedure is applied recursively to each square, to infinity.
4. The figure 10 shows the three iterations of the PYTHAGORE Tree Structure.

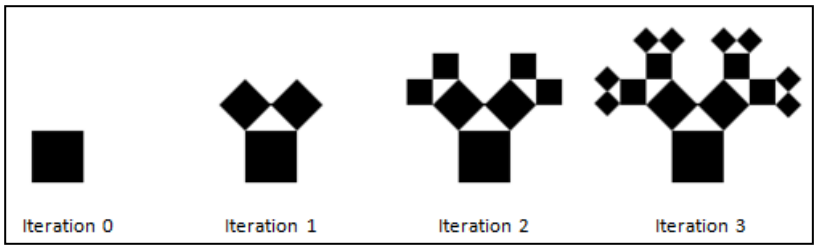


Figure 10: the three first iterations of the PYTHAGORE Tree Structure

The HAUSDORFF dimension for the PYTHAGORE Tree Structure is given by the equation (10).

$$d = \frac{\ln(n)}{\ln(h)} = \frac{\ln(2)}{\ln(\frac{2}{\sqrt{2}})} = 2 \tag{10}$$

4.5 The circular structure "APOLLONIUS circle"

This structure was invented by the Greek mathematician APOLLONIUS of Perga. Apollonius circles are tangent to one over other.

For the construction of this geometry the following steps are followed:

1. We begin with three circles C1, C2 and C3 of any size, each of which is tangent to the others.
2. According to Apollonius, there are two other circles which do not intersect, C4 and C5, which have the property of being tangent with the original three circles - these were called Apollonius circles.
3. To construct the other circles, each time we take 3 tangent circles and we repeat the step 2.

The Figure 11 shows some iteration of the APOLLONIUS Circles.

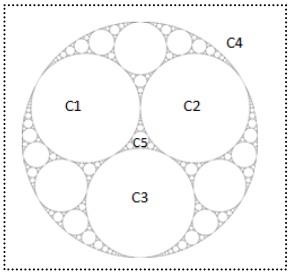


Figure 11: APOLLONIUS Circles

The HAUSDORFF dimension for the Apollonius Circles Structure is 1.3 [6].

4.6 CANTOR Set

CANTOR Set was invented by the German mathematician Georg CANTOR. It is built iteratively from the segment [0, T] by removing a central portion (a third for example); then the operation is repeated on the remaining two segments, and so on. The figure 12 shows the 6 first iterations of the Cantor set structure.

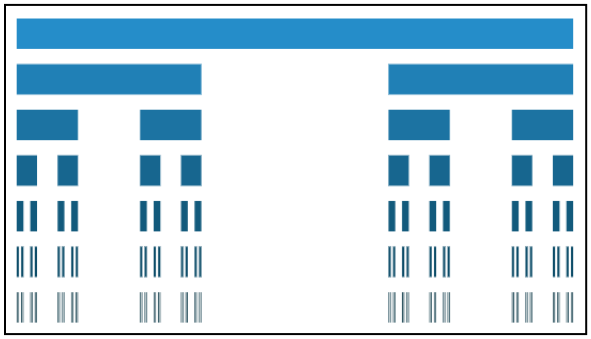


Figure 12: the 6 first iterations of the CANTOR SET structure

The HAUSDORFF dimension of the CANTOR Set fractal is given by the equation (11):

$$d = \frac{\ln(n)}{\ln(h)} = \frac{\ln(2)}{\ln(\alpha)} \tag{11}$$

Where $1/\alpha$ is the remaining part of the initial length of the first iteration.

In the example (figure 12), $\alpha = 3$.

So the HAUSDORFF dimension is given by the equation (10).

$$d = \frac{\ln(n)}{\ln(h)} = \frac{\ln(2)}{\ln(3)} = 0.63 \tag{12}$$

4.7 The HILBERT curve

The Hilbert curve is described for the first time by the German mathematician David Hilbert in 1891 [7]. The Hausdorff dimension is 2 and the method of construction is described in Figure 13 [8].

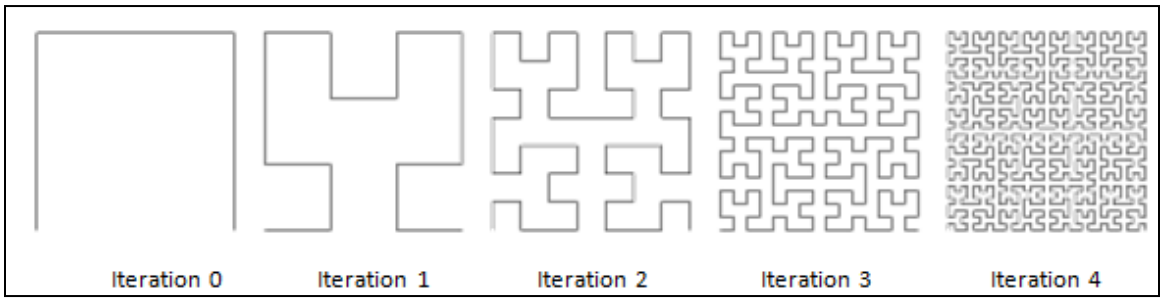


Figure 13: the first four iterations of the HILBERT Curve Structure

4.8 The MINKOWSKI Curve

This curve was invented by the German mathematician Hermann MINKOWSKI. The generation of this fractal is described in Figure 14.

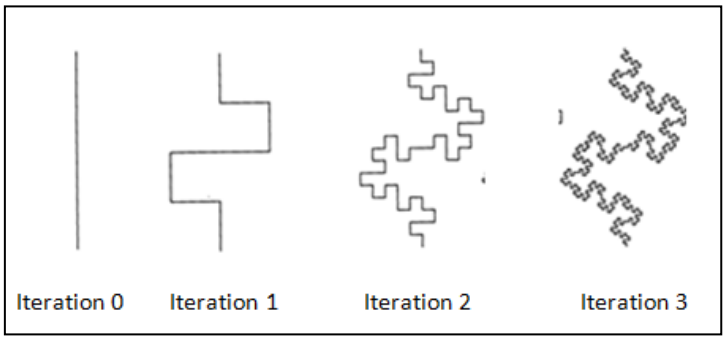


Figure 14: the first three iterations of the MINKOWSKI Curve Structure

The HAUSDORFF dimension of the MINKOWSKI curve fractal is given by the equation 13.

$$d = \frac{\ln(n)}{\ln(h)} = \frac{\ln(8)}{\ln(4)} = 1.5 \tag{13}$$

5 Fractal Antennas Geometries

5.1 Background of the study and issues

With the proliferation and miniaturization of telecommunications systems and their integration in restricted environments, such as Smart-phones, tablets, cars, airplanes, and other embedded systems. The design of compact multi-bands and Ultra Wide Band (UWB) antennas becomes a necessity.

For designing this kind of antennas, two techniques are used:

a) Designing multi-bands antennas operating in several frequencies bands. Several studies have been made to design this kind of antennas by using fractal geometries or adding slots to the radiating elements [9]-[14].
b) Designing UWB antennas operating in the frequencies bands exceeding 500MHz or having a fractional bandwidth of at least 0.20, UWB wireless communication occupies a bandwidth from 3.1 to 10.6 GHz (based on the FCC "Federal Communication Commission") [14]-[22].

According to our literature searches, the use of fractal antennas allows to have multi-band and Broad-Band behavior.

5.2 The use of the fractal antennas

Several fractal geometries have been adopted to design fractal antennas including: the KOCH curve, the KOCH snowflake, the SIERPINSKI triangle, the SIERPINSKI carpet and other structures.

5.2.1 The KOCH structures

- ***Wire antennas***

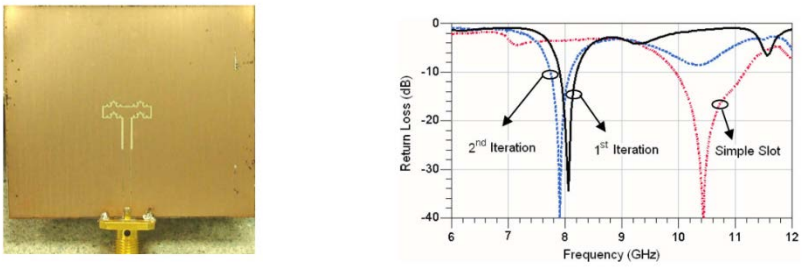
The KOCH curve was used for the design of wire antennas with good performances. In 2000, PUENTE demonstrated that we can improve the efficiency of the antennas by increasing the number of iterations [23]. In 2002, BEST confirmed that the resonance frequency and the radiation resistance decrease when

we increase the number of iterations while maintaining a fixed length of the structure [24]. In 2003 and 2004, VINOY demonstrated that the dimension of the fractal geometry affects the resonance frequencies. For each dimension, the relationships between the resonance frequencies are unchanged even if the number of iterations increases [25-26].

- ***Patch antennas***

The KOCH curve was used also for the design of patch antennas. In fact, it was used for the construction of the radiating elements, or for the setup of the slots on the radiating elements, or for the design of the ground planes.

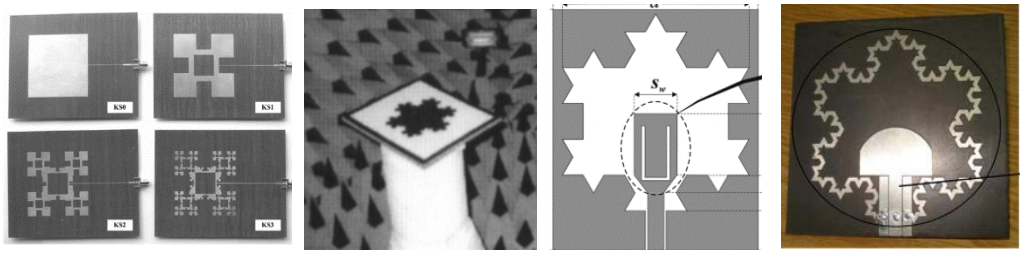
In 2007, SUNDARAM used the KOCH curve to design slots on a patch antenna, increasing the number of iterations; we decrease the resonant frequencies without increasing the size of the patch. This technique is a way to miniaturize the antennas. The figure 15 shows the structure of the proposed antenna and its measured return loss [27].



(a) The structure of the proposed antenna **(b) The measured return loss**

Figure 15: the structure and the measured return loss of the proposed antenna by SUNDRAM [27]

In 2008, CHEN designed a patch antenna using the KOCH and SIERPINSKI structures with a lower resonant frequency and a largest bandwidth compared to the traditional patch antenna [28]. Also, ANGUERA presented the performances of a patch antenna based on a KOCH fractal structure [29]. KRISHNA and PATNAM used the KOCH structures in the CPW-fed patch ground plane antennas and have increased the bandwidth and decreased the resonant frequencies of the antennas [30]-[31]. The figure 16 shows the proposed antenna by CHEN, ANGUERA, KRISHNA and PATNAM.

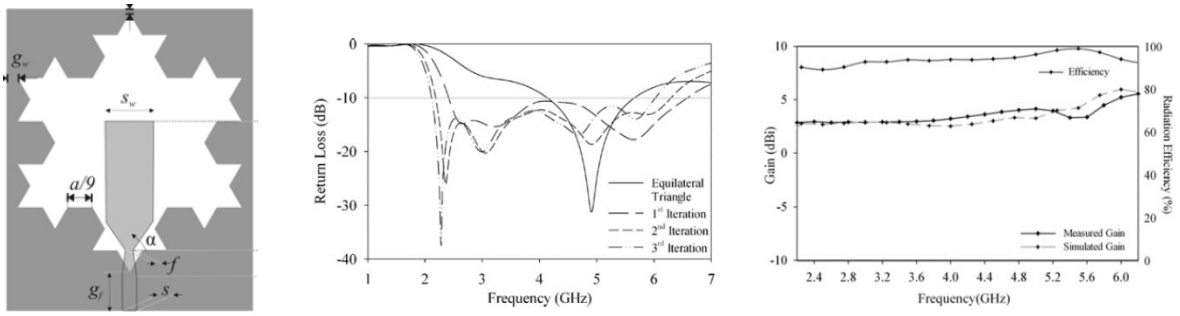


(a) **(b)** **(c)** **(d)**

Figure 16: the structures proposed by (a) CHEN (b) ANGUERA (c) KRISHNA (d) PATNAM [28]-[32]

In 2009, KRISHNA has designed another patch antenna having a KOCH structure on the ground plane with an Ultra Wide Bandwidth and a high gain. The figure 17 shows the structure of the proposed

antenna, the return loss versus the frequency and versus the iterations number, the gain and the efficiency versus the frequency [32].



(a) (b) (c)

Figure 17: the structures proposed by KRICHNA (a), the return loss versus the iteration number and versus the frequency (b), the gain and the efficiency of the proposed antenna (c) [32]

In 2010, BIN YOUNAS presented a patch antenna with a high directivity. The resonant element and a slot are designed with the KOCH curve [33]. YUSOP used the KOCH structure for the design of an antenna fed by induction and has achieved an important gain [34]. The figure 18 shows the structures proposed by BIN YOUNAS and YUSOP.

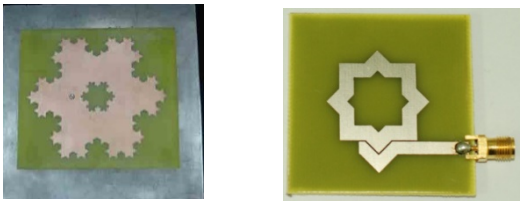
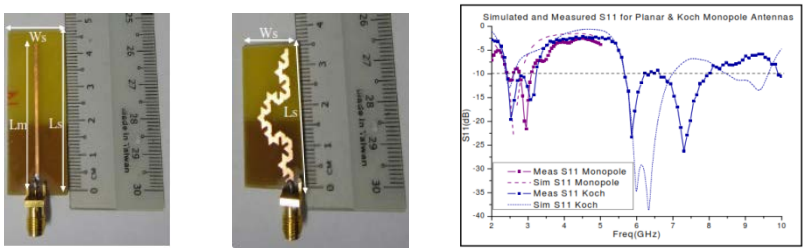


Figure 18: the structures proposed by BIN YOUNAS (a), and by YUSOP (b) [33]-[34]

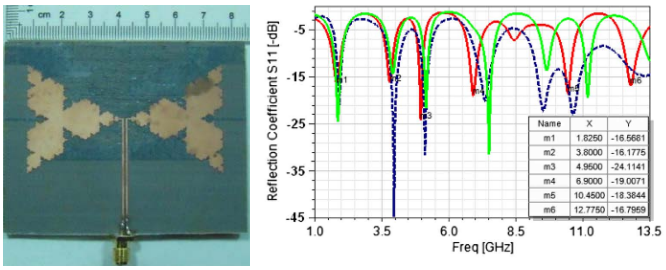
In 2011, ISMAHAYATI presented a comparative study between the performances of a simple monopole antennas and a KOCH monopole structure. The last one presented a multi-band and broad-band behavior [35]. The figure 19 shows the two structures and the comparison of the return Loss versus the frequency for the two structures.



(a) (b) (c)

Figure 19: the structures proposed by ISMAHAYATI, (a) Simple Monopole (b) KOCH Monopole, and (c) the comparison of the return Loss versus the frequency for the two structures [35]

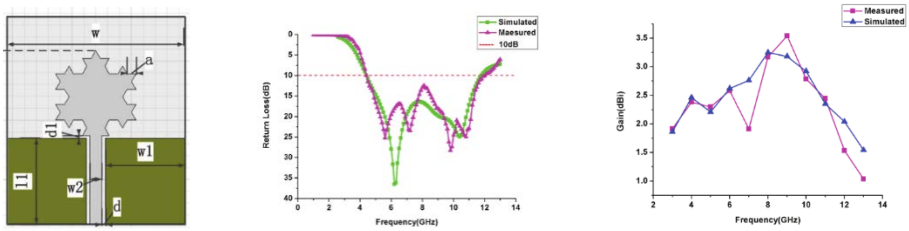
In 2012, DAOTIE LI combined the two structures BOW-TIE and KOCH. This antenna presents a multi-band and wide-band behavior [36]. The figure 20 shows the structure of the proposed antenna and the return loss of the antenna versus the frequency versus the iterations number.



(a) (b)

Figure 20: the structure proposed by DAOTIE LI (a) and its return loss versus the frequency and versus the iterations number [36]

Also, DONG LI presented a miniaturized patch antenna as a structure KOCH 2nd iteration with significant gains and a wide bandwidth [37]. The figure 21 shows the structure proposed by DAOTI LI, the simulated and measured return loss, and the simulated and measured gain.

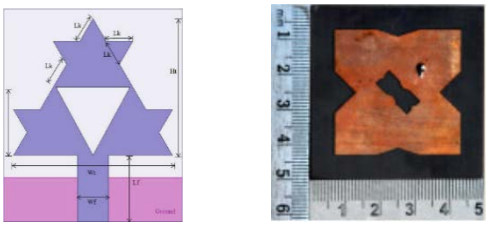


(a) (b) (c)

Figure 21: the structure proposed by DAOTIE LI (a), the simulated and measured return loss versus the frequency (b), and the simulated and measured gain [37]

In 2013, JEEMON combined the two structures KOCH and SIERPINSKI to design a miniaturized antenna for UWB applications and for WIMAX and WLAN applications [38].

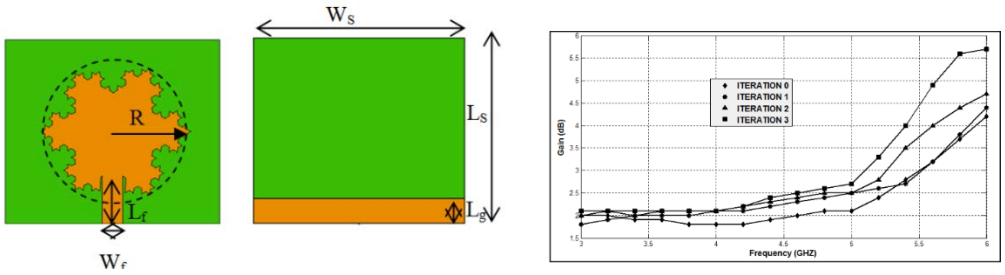
In 2014, REDDY designed a patch antenna with a first iteration KOCH curve borders. This structure performed a high gain and it is operational for WIMAX and WLAN applications [39]. The figure 22 shows the structures proposed by JEMON and REDDY.



(a) (b)

Figure 22: the proposed antennas by JEEMON (a), and REDDY (b) [37]-[39]

Also REHA studied the behavior of a CPW-Fed KOCH SNOWFLAKE Fractal Antenna for UWB Wireless Applications and demonstrated that increasing the number of iterations allows obtaining a low profile antenna with a high gain, a multi-band and a broad-band behavior [14]. The figure 23 shows the structure studied and the gain of the antenna versus the frequency and the iterations number.



(a) (b)

Figure 23: the antennas studied by REHA (a), and the simulated gain versus the frequency and the iterations number (b) [14]

5.2.2 The SIERPINSKI structures

- ***SIERPINSKI triangles***

In 1998, BORJA has implemented a patch antenna using SIERPINSKI triangle network and has obtained a multi-band antenna by incorporating more iteration [40]. Also, PUENTE has modeled the behavior of this kind of fractal antennas [41].

In 1999 GONZALEZ studied experimentally the distribution of currents in a SIERPINSKI fractal antenna and demonstrated that this distribution follows the principle of self-similarity [42]. In 2000, he predicted this kind of antenna when the flare angle is modified [43]. The figure 24 shows the structure studied by BORJA, PUNTE, and GONZALEZ.

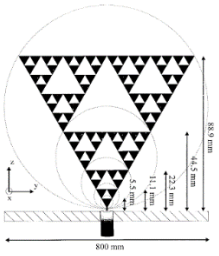
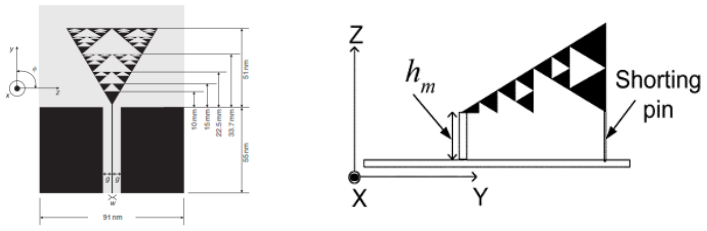


Figure 23: the antenna studied by BORJA, PUNTE, and GONZALEZ [41]

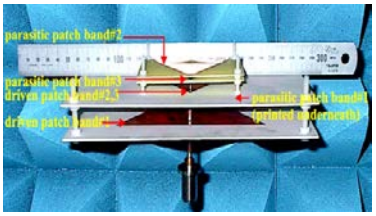
In 2004, KITLINSKI confirmed the results found previously and improved the behavior of the SIERPINSKI fractal antenna by adopting the CPW-feeding technique [44]. SONG studied the behavior of these antennas by leaving only half of the antenna and adding short circuits. This technique allows miniaturization while maintaining the same performance of the complete antennas [45]. The figure 24 shows the antennas studied by KITLINSKI and SONG.

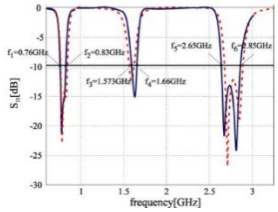


(a) (b)

Figure 24: the antennas studied by KITLINSKI (a), and SONG (b) [44]-[45]

In 2006, ANGUERA set up the SIERPINSKI fractal antenna on several layers of substrate and designed a multi-band antenna operating for all mobile networks 2G-4G with good performances [46]. The figure 25 shows the proposed structure and its return loss.

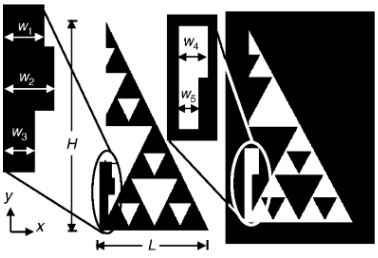


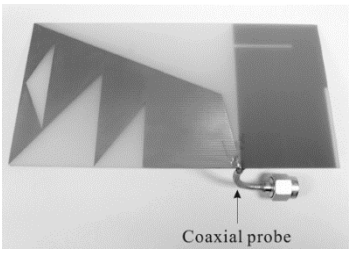


(a) (b)

Figure 25: the structure proposed by ANGUERA (a) and its return loss versus the frequency (b) [46]

In 2007, HWANG and VEMAGRI have developed semi-SIERPINSKI fractal antennas respectively for UHF-RFID (Ultra High Frequency - Radio Frequency Identification) and mobile networks 2G-4G [47-48]. The figure 26 shows the structures proposed by **HWANG and VEMAGRI** .





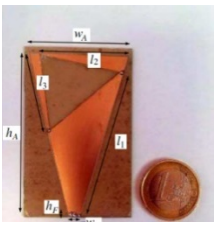
(a) (b)

Figure 26: the structures proposed by HWANG (a) and VEMAGRI (b) [47]-[48]

In 2009, KRZYSZTOfiK presented a modified and miniaturized SIERPINSKI Fractal structure to cover the two ISM (Industrial, Scientific and Medical) bands (2.4 and 5.2 GHz) with better performance compared to traditional structures [49].

In 2011, SOH has setup the SIRPINSKI structure in a PIFA (Planar Inverted F Antenna) for IEEE 802.15 WPAN LR (Low Rate Wireless Personnal Area Network) applications [50].

In 2013, VIANI has changed the basic structure of the SIERPINSKI geometry and has designed an antenna with high gains for LTE and WiMAX applications [51]. The figure 27 shows the structures proposed by **KRZYSZTOfiK, SOH and VIANI**.



(a) (b) (c)

Figure 27: the structures proposed by KRZYSZTOfiK (a), SOH (b), and VIANI [49]-[51]

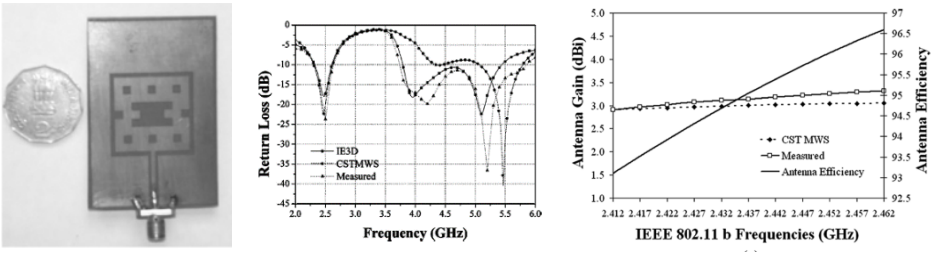
- ***SIERPINSKI Carpet***

In 1999, Werner showed the effect of the iterations number on the behavior of fractal antennas in general and in particular of the SIERPINSKI Carpet fractal structure [52].

In 2004, Ban-Leong OOI proposed a modified SIERPINSKI carpet structure to improve the bandwidth and the gain compared to a standard SIERPINSKI structure [53].

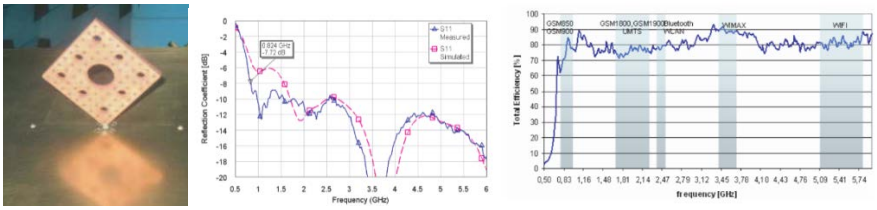
In 2008, GHATAK proposed other changes to the standard SIERPINSKI carpet structure to adapt it to wireless networks IEEE 802.11a / b WLAN and HyperLAN2 applications with a good efficiency [54]. The figure 28 shows the proposed structure, its return loss, its gain and efficiency versus the frequency.

In 2010, ANGUERA has designed a 3D SIERPINSKI carpet antenna for WiFi, 2G, 3G, WIMAX, and Bluetooth applications, with very high gains and good efficiency [55]. The figure 29 shows the proposed structure, its return loss, and its efficiency versus the frequency.



(a) (b) (c)

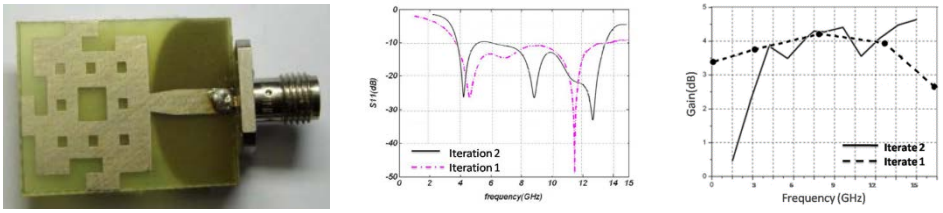
Figure 28: the structure proposed by GHATAK (a), its return loss (b), its gain and efficiency versus the frequency(c) [54]



(a) (b) (c)

Figure 29: the structure proposed by ANGUERA (a), its return loss (b), and its efficiency versus the frequency (c) [55]

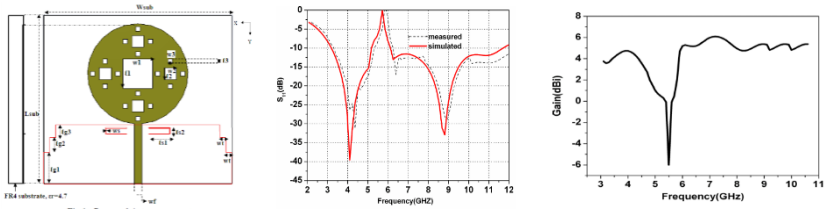
In 2011, ORAIZI combined the SIERPINSKI Carpet structure with the GIUSEPE PEANO structure. The result is a miniaturized Ultra Wide Band antenna, with high efficiency and high gain [56]. The figure 30 shows the proposed structure, its return loss, and its gain versus the frequency.



(a) (b) (c)

Figure 30: the structure proposed by ORAIZI (a), its return loss (b), and its gain versus the frequency (c) [56]

In 2013, BISWAS has set up the SIERPINSKI Carpet structure on the circular patch antenna and fed by a microstrip line and having slots on the ground plane. This new structure allowed having an Ultra Wideband behavior with significant gains [57]. The figure 31 shows the proposed structure, its return loss, and its gain versus the frequency.



(a) (b) (c)

Figure 31: the structure proposed by BISWAS (a), its return loss (b), and its gain versus the frequency (c) [57]

5.2.3 The Tree Structure

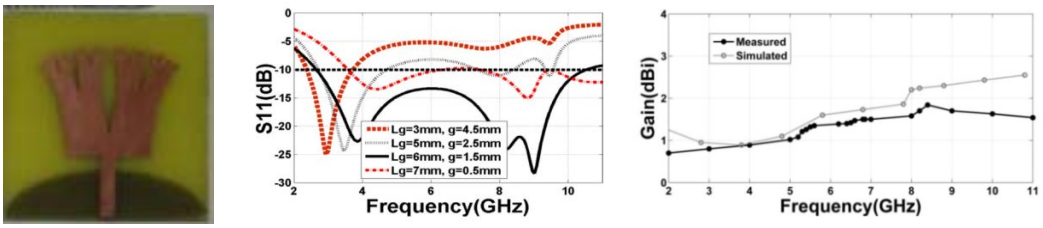
In 1986, according to our literature searches, KIM has published the first paper on the fractal antennas and especially on the random tree fractal antennas [58].

In 1999, WERNER studied the effect of iterations number on the behavior of fractal antennas in general and in particular of the Tree structures [59].

In 2004, PETKO studied the 3D TREE Fractal antennas and found out the relationship between the geometry parameters and the performances of this kind of fractal structures [60].

In 2009, He studied the array of TREE Fractal antennas and he demonstrated that we can have better results with this technique [61].

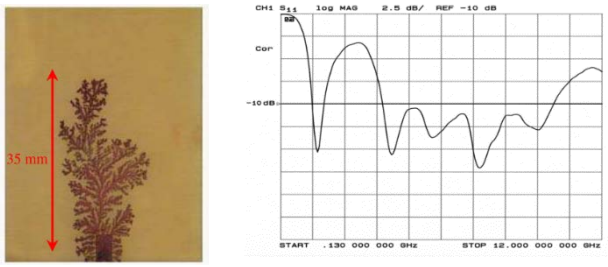
In 2011, POURAHMADAZA created a modified patch antennas fed by microstrip lines and having the shape of "PYTHAGORE TREE". He demonstrated that by increasing the number of iterations we can have miniaturized antennas with more resonance frequencies, an Ultra Wideband behavior, good efficiency and better matching [62]. The figure 32 shows the proposed structure, its return loss, and its gain versus the frequency.



(a) (b) (c)

Figure 32: the structure proposed by POURAHMADAZA (a), its return loss (b), and its gain versus the frequency (c) [62]

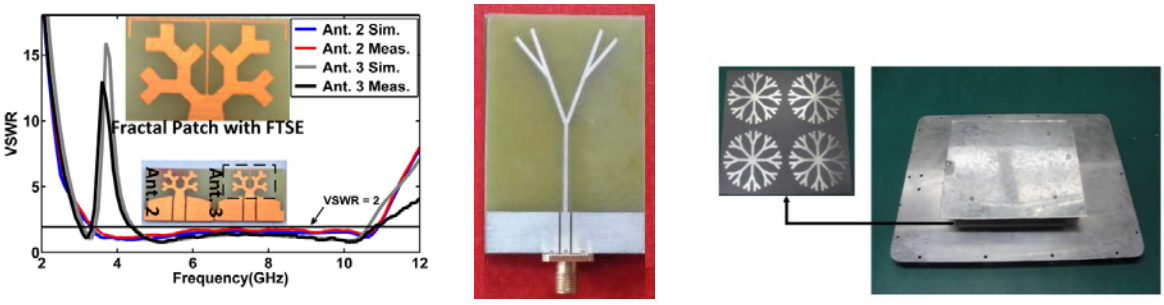
In 2013, DUMOND studied experimentally the Random 2D-TREE fractal antennas and he demonstrated that with this kind of structures, we can have antennas with good performances, multi-band and UWB behavior [63].



(a) (b)

Figure 33: the structure proposed by DUMOND (a) and its return loss (b) [63]

Also, NASER-MOGHADASI, has introduced a miniaturized TREE fractal structure operational for all the UWB applications, he set up also parasite elements in order to eliminate some frequencies bands [64]. VARADHAN proposed a Tri-bands TREE fractal antenna for The RFID applications [65]. LIU proposed another structure with a high gain for the UHF-RFID, this antenna is composed of a traditional patch and another layer composed of a four TREE fractal elements [66]. The figure 34 shows the proposed structures by -MOGHADASI, VARADHAN, and LIU.



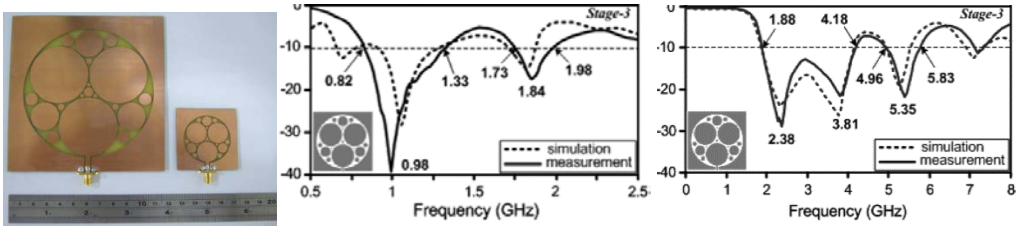
(a) (b) (c)

Figure 34: the proposed structures by NASER-MOGHADASI (a) , VARADHAN (b), and LIU (c) [64]-[65]

5.2.4 The circular Structures

- ***Apollonius circle***

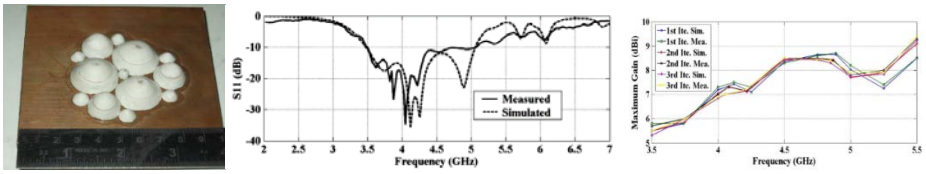
In 2008, CHANG proposed two Apollonius circles antennas with very high gains and a multi-band behavior [67]. The figure 35 shows the proposed structures and its return loss for the λ/2 and λ/4 design.



(a) (b) (c)

Figure 35: the proposed structures by CHANG (a), its return loss for the λ/2 design (b), and λ/4 design (c) [67]

In 2014, MUKHERJEE has set up a 3D Apollonius circles with high gain and a broad-band behavior [68]. The figure 36 shows the proposed structure, its return loss, and its gain.



(a) (b) (c)

Figure 36: the proposed structure by MUKHERJEE (a), its return loss (b), and its gain (c) [68]

- ***Other circular fractal structures***

Several studies have been made on the circular fractal geometries without being as Apollonius circles. In 2006, DING has designed a circular fractal antenna fed by a microstrip line, combining circles and triangles as shown in Figure 37. The proposed antenna is a small structure, broad-band and can be used in UWB applications [69].

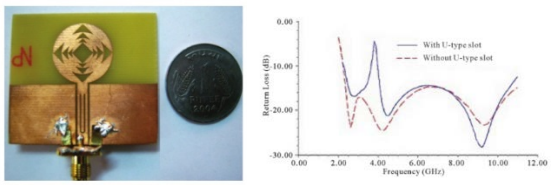
(a) (b)

Figure 37: the circular fractal structures proposed by DING (a) and its the return loss (b) [69]

In 2008, as shown in Figure 38, KUMAR proposed the same modified structure. The proposed structure is a miniaturized antenna with multi-band behavior and a high gain [70].

Figure 38: the circular fractal structure proposed by KUMAR [70]

In 2010, Kumar proposed another circular fractal antenna (Figure 39). The proposed structure provides a broad-band behavior. The setup of the slots on the microstrip line allows the reject of some frequencies bands [71]. The figure 39 shows the proposed structure and its return loss with and without the slot.



(a) (b)

Figure 39: the circular fractal structure proposed by KUMAR [71]

In 2012 MOHAMMADSHAH has proposed the circular fractal structure shown in the Figure 40. The proposed structure is a broad-band antenna, operational from 2 to 21GHz with a good efficiency [72].

Figure 40: the circular fractal structure proposed by MOHAMMADSHAH [72]

In 2014, REDDY has studied the effect of the introduction of the semicircular fractals on the edges of a patch antenna as shown in Figure 41. The proposed structure provides important gains for WiFi, WiMAX and WLAN applications [73].

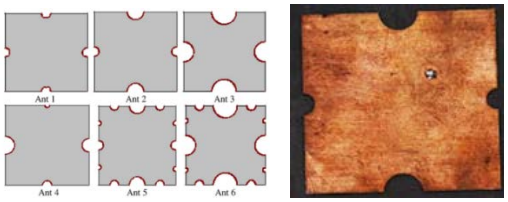
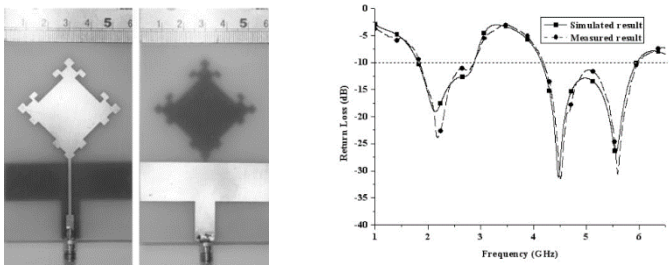


Figure 41: the circular fractal structure proposed by REDDY [73]

5.2.5 The MINKOWSKI curve

In 2002 and 2003, GIANVITTORIO and BEST have studied the effect of iterations number on the resonance frequencies number. They showed that for the MINKOWSKI curve wire antennas, the number of resonant frequencies increases by increasing the number of iterations [74-75].

In 2009, MAHATTHANAJATUPHAT has designed a modified patch antenna with MINKOWSKI Curve on the borders of the radiating element for UMTS, WLAN and WIMAX applications [76]. The figure 42 shows the proposed structure and its return loss.



(a) **(b)**

Figure 42: the structure proposed by MAHATTHANAJATUPHAT (a) and its return loss (b) [76]

In 2011, CHEN has designed a modified patch antenna by MINSOWSKI curve to reduce the RCS (Radar Cross Section) [77]. Also, MORAES has studied the difference between a square patch and a triangular patch antenna. The antennas have been optimized to have initially the same resonant frequency. The introduction of MINKOWSKI curve contour allows having lower resonance frequencies compared to the introduction of KOCH curve contour [78].

In 2012, NAJI has studied a patch antenna with Proximity-coupled feed, modified by a lot of MINKOWSKI curve contour configurations. The proposed structures are very miniaturized, with an important gain, and operational for 5.8GHz-RFID applications [79-80].

In 2013 and 2014, LATA, DHAR, AHMAD, SINGH AND ANCY have proposed patch antennas with modified contours with a MINKOWSKI curve. These structures have a Multi-band and a broad-band behavior [81]-[85]. The figure 43 shows the structures proposed.

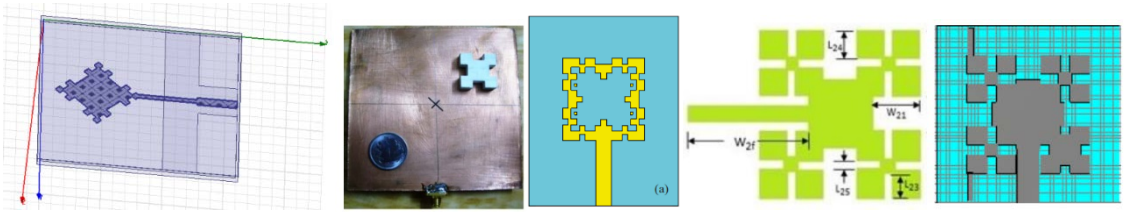


Figure 43: the structures proposed by LATA (a), DHAR (b), AHMAD (c), SINGH (d), and ANCY (e) [81]-[85]

5.2.6 The HILBERT Curve

In 2003, GONZALEZ-ARBESÚ has studied the parameters and the behavior of 2D and 3D HILBERT curve wire antennas [86].

In 2006, MURAD has developed a modified patch antenna using the Hilbert curve. The proposed antenna is a miniaturized structure for the 2.45GHz RFID applications [87].

In 2010, SANZ has studied the difference between the spiral wire antennas and the wire antennas based on the Hilbert geometries. He has showed that this second structure provides a higher bandwidth [88]. The figure 44 shows the studied structures and its VSWR (Voltage Standing Wave Ratio).

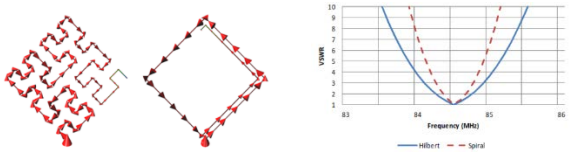
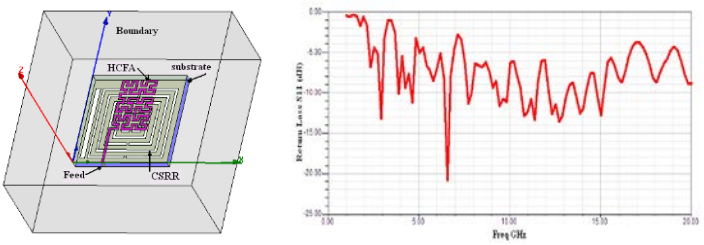


Figure 44: the studied structures by SANZ (a) and its VSWR (b) [88]

HUANG has set up a PIFA antenna based on a HILBERT structure, operational for the 2.4GHz applications [89]. The figure 45 shows the proposed structure and its return loss.

Figure 45: the proposed structure by HUANG (a) and its Return Loss (b) [89]

In 2012, SUGANTHI has studied some modified patch antennas based on the HILBERT structure. The proposed antennas are operational for medical applications [90]. The figure 46 shows the proposed structure and its return loss.

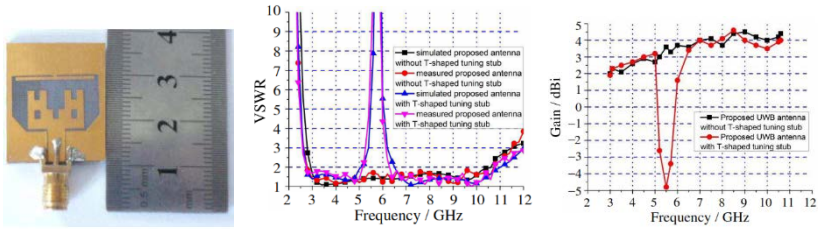


(a) (b)

Figure 46: the proposed structure by SUGANTHI (a) and its Return Loss (b) [90]

5.2.7 CANTOR Set

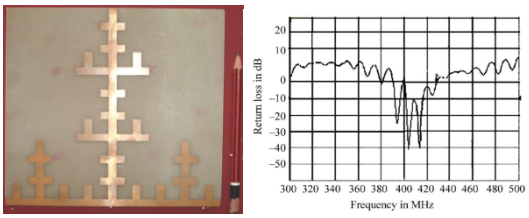
In 2011, LI has designed a miniaturized and modified patch antenna fed by a microstrip line having a radiating element in the form of a CANTOR set. This structure is operational for UWB applications and having an important gain except for the 5 - 6.3GHz.applications [91]. The figure 47 shows the proposed structure, its VSWR, and its gain.



(a) (b) (c)

Figure 47: the proposed structure by LI (a), its VSWR (b), and its gain (c) [91]

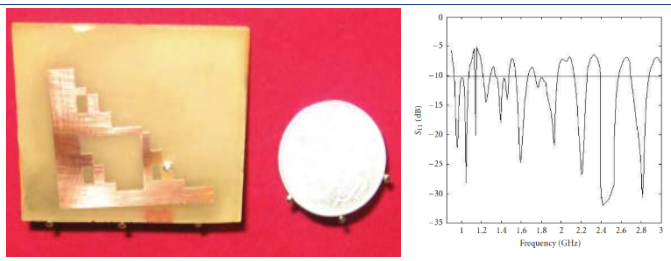
Also, SRIVATSUN has proposed the modified CANTOR set structure with for MICS (Medical Implant Communication Service) applications [92]. The figure 48 shows the proposed structure, and its return loss.



(a) (b)

Figure 48: the proposed structure by SRIVATSUN (a), and its return loss (b) [92]

In 2012, SRIVATSUN has the same structure for the IEEE 802.11b WLAN applications, 802.15, PCS, GSM, DCS, IMT, UMTS, Wi-Fi, and WLAN with good performances [93]. The figure 49 shows the proposed structure and its return loss.



(a) **(b)**

Figure 49: the proposed structure by SRIVATSUN (a), and its return loss (b) [93]

6 Conclusion

In this paper, some generalities of fractal geometries and their dimensions are presented. We have described also some linear fractal geometries such as KOCH, SIERPINSKI, DRAGON, TREE, CIRCULAR, CANTOR SET, HILBERT, MINKOWSKI, and we have discussed the applications of these geometries and their performances in the design of the miniaturized antennas.

The common feature of all the presented works is that the fractal antennas allow a Broad-Band and a Multi-band behavior. Also, the fractal antenna allows having miniaturized antennas with good performances.

In fact, for the KOCH fractal structures, the wire antennas allows a multi-bands behavior, the combination with the patch antennas decreases the resonant frequencies and the introduction of the CPW feeding allows a Broad-Band behavior. For the SIERPINSKI fractal structures, the patch antennas allows having a multi-band behavior, also, the introduction of the CPW feeding allows a broad-Band behavior. For the TREE fractal structures, we can have multi-band antennas. The use of PYTHAGORE TREE structures allows having broad-band antennas. The Circular fractal structures allows having a multi-band behavior, the use of 3D structures allows having a broad-band antennas with a high gains. The use of MINKOWSKI fractal structures, allows having a multi-band and broad-band antennas. For the HILBERT fractal structures, we can have multi-band antennas. For the CANTOR set fractal structures, we can design broad-band antennas.

According to our literature search, it is clear that some have been extensively used such as KOCH, SIERPINSKI and TREE structures, other structures were rarely used such as CIRCULAR structures, CANTOR Set, but some of them weren't studied such as DRAGON, H-TREE structures. Also, several studies have combined two fractal structures or have modified the original fractal structures.

In the next works, we will try to study the fractal structures that have never been studied such as DRAGON and H-Tree structures, and we will try also to study the performances of the fractal antennas on the networking configurations.

REFERENCES

[1]. Mandelbrot, B.B., "the Fractal Geometry of Nature", W.H. Freeman and Company, New York, 1983.

[2]. MANDELBROT, B.B, « Les Objets Fractals ». 4e éd., Flammarion, 1995.

[3]. Jacques Dubois, Jean Chaline, Claude Allègre, « Le Monde Des Fractales : La Géométrie Cachée De La Nature », ellipse, 2006.

[4]. Constantine A. Balanis, "ANTENNA THEORY - ANALYSIS AND DESIGN", A JOHN WILEY & SONS, INC. PUBLICATION, 2005.

[5]. Kenneth FALCONER, "Fractal Geomerty : Matimatical Foundation and applications", A JOHN WILEY & SONS, INC. PUBLICATION, 2003

[6]. Curtis T. McMullen, "Hausdorff dimension and conformal dynamics III: Computation of dimension", Harvard University, 1 Oxford St, Cambridge, pp 17, October, 1997.

[7]. D. Hilbert, « Über die stetige Abbildung einer Linie auf ein Flächenstück », Math. Ann., vol. 38, p. 459-460, 1891.

[8]. Peitgen, H.-O. and Saupe, D. (Eds.). « The Science of Fractal Images ». New York: Springer-Verlag, pp. 278 and 284, 1988.

[9]. A. Elouadih, A. Oulad-Said and M. Hassani, "Design and Simulation of a PIFA Antenna for the Use in 4G Mobile Telecommunications Networks," Int'l J. of Communications, Network and System Sciences, Vol. 6 No. 7, 2013, pp. 325-332. doi: 10.4236/ijcns.2013.67035.

[10]. C. Varadhan, J. K. Pakkathillam, M. Kanagasabai, R. Sivasamy, R. Natarajan and S. K. Palaniswamy, "Triband Antenna Structures for RFID Systems Deploying Fractal Geometry," IEEE Antennas and Wireless Propagation Letters, Vol. 12, 2013, pp. 437-440.

[11]. D. C. Chang, B. H. Zeng, and J. Liu, "CPW-fed circular fractal slot antenna design for dual-band applications," IEEE Trans. Antennas Propag., vol. 56, no. 12, pp. 3630–3637, Dec. 2008.

[12]. Abdelati REHA, Ahmed OULAD SAID, "Tri-Band Fractal Antennas for RFID Applications," Wireless Engineering and Technology, Vol. 4 No. 4, pp. 171-176, 2013.

[13]. Abdelati REHA, Marouane BOUCHOUIRBAT, « A Dual-Band Rectangular CPW Folded Slot Antenna for GNSS Applications » ,International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 3, Issue 8, pp 11055- 11061, August 2014.

[14]. Abdelati REHA, Abdelkebir EL AMRI, Othmane BENHMAMMOUCH, Ahmed OULAD SAID, "CPW-Fed KOCH SNOWFLAKE Fractal Antenna for UWB Wireless Applications" , Transaction on Networks and Communications, Vol. 2, Issue 4, pp 38-53, August 2014.

[15]. Guo-Ping Gao, Bin Hu, and Jin-Sheng Zhang, Design of a Miniaturization Printed Circular-Slot UWB Antenna by the Half-Cutting Method, IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS, VOL. 12, pp 567-570, 2013

[16]. Takumi Sugitani, Shinichi Kubota, Akihiro Toya, Xia Xiao, and Takamaro Kikk awa, A Compact 4x4 Planar UWB Antenna Array for 3-D Breast Cancer Detection, IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS, VOL. 12, pp 733-736,2013.

[17]. P. Thriuvalar Selvan, S. Raghavan, M. Gopinath, CPW-Fed Folded-UWB Monopole Slot Antenna for WiMAX and WLAN Applications, IEEE ICECCN, pp 696-699,2013.

[18]. Xue Ni Low, Zhi Ning Chen, Terence S. P. See, A UWB Dipole Antenna With Enhanced Impedance and Gain Performance, IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, VOL. 57, NO. 10, pp 2959-2966, OCTOBER 2009.

[19]. A. Shaker, S. Zainud-Deen, K. Mahmoud and S. Ibrahem, "Compact Bluetooth/UWB Antenna with Multi-Band Notched Characteristics," Journal of Electromagnetic Analysis and Applications, Vol. 3 No. 12, pp. 512-518, 2011.

[20]. A. Hosseinzadeh, M. Mirmozafari, M. Varnoosfaderani, C. Ghobadi and J. Nourinia, "A U-Shaped UWB Antenna with Band-Notched Performance," Wireless Engineering and Technology, Vol. 4 No. 4, pp. 177-180, 2013.

[21]. G. TATSIS, V. RAPTIS and P. KOSTARAKIS, "Design and Measurements of Ultra-Wideband Antenna," Int'l J. of Communications, Network and System Sciences, Vol. 3 No. 2, pp. 116-118, 2010.

[22]. M. Jahanbakht and A. Neyestanak, "A Survey on Recent Approaches in the Design of Band Notching UWB Antennas," Journal of Electromagnetic Analysis and Applications, Vol. 4 No. 2, pp. 77-84, 2012

[23]. Carles Puente Baliarda, Jordi Romeu, and Angel Cardama, « The Koch Monopole: A Small Fractal Antenna", IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, VOL. 48, NO. 11, pp 1773-1781, NOVEMBER 2000.

[24]. Steven R. Best , "On the Resonant Properties of the Koch Fractal and Other Wire Monopole Antennas", IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS, VOL. 1, pp 74-76, 2002

[25]. K. J. Vinoy, Jose K. Abraham, and Vijay K. Varadan, "On the Relationship between Fractal Dimension and the Performance of Multi-Resonant Dipole Antennas Using Koch Curves", IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, VOL. 51, NO. 9, pp 2296-2303, SEPTEMBER 2003.

[26]. K. J. Vinoy, jose k. Abraham, and v. K. Varadan, "impact of fractal dimension In the design of multi-resonant fractal antennas", Fractals , Vol. 12, No. 1, pp 55–66, 2004.

[27]. Ananth Sundaram, Madhurima Maddela, , and Ramesh Ramadoss, "Koch-Fractal Folded-Slot Antenna Characteristics", IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS, VOL. 6, pp 219-222, 2007.

[28]. Wen-Ling Chen, Guang-Ming Wang, and Chen-Xin Zhang, "Small-Size Microstrip Patch Antennas Combining Koch and Sierpinski Fractal-Shapes", IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS, VOL. 7, pp 738-741, 2008

[29]. Jume Anguera et Al, "Metallized Foams for Fractal-Shaped Microstrip Antennas" IEEE Antennas and Propagation MagaZine, Vol. 50, No.6, pp 20-38, December 2008.

[30]. Deepti Das Krishna et Al, "CPW-Fed Koch Fractal Slot Antenna for WLAN/WiMAX Applications", IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS, VOL. 7, pp 389-392, 2008.

[31]. Rao Hanumantha Patnam, “Broadband CPW-Fed Planar Koch Fractal Loop Antenna”, IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS, VOL. 7, pp 429-431, 2008.

[32]. D.D. Krishna et Al, "Compact wideband Koch fractal printed slot antenna", IET Microw. Antennas Propag, Vol. 3, Iss. 5, pp. 782–789, 2009

[33]. Abbas Bin Younas, Zubair Ahmed, Mojeeb Bin Ihsan, "A New High-Directivity Fractal Antenna Based on the Modified Koch Snowflake Geometry", Asia-Pacific Microwave Conference 2010, pp 191-194, 2010

[34]. M. A. M. Yusop, M. K. A. Rahim, M. F. Ismail, A. Wahid , "Circular Polarization Fractal Koch Microstrip Patch Antenna using Single-fed EM Coupled Ring Resonators ", IEEE Asia-Pacific Conference on Applied Electromagnetics (APACE 2010) , pp 1-4, 2010

[35]. A. Ismahayati, P.J Soh, R.Hadibah, G.A.E Vandenbosch , "Design and Analysis of a Multiband Koch Fractal Monopole Antenna", IEEE International RF and Microwave Conference (RFM 2011), 12th - 14th December 2011, Seremban, Malaysia, pp 58-62, 2011

[36]. Daotie Li, and Jun-fa Mao, "A Koch-Like Sided Fractal Bow-Tie Dipole Antenna", IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, VOL. 60, NO. 5, pp 2242-2251, MAY 2012.

[37]. Dong Li, Fu-Shun Zhang , Zong-Ning Zhao, Liu-Tao Ma, and Xu-Nan Li , "UWB antenna design using patch antenna with Koch fractal boundary", IEEE International Conference on Microwave and Millimeter Wave Technology (ICMMT 2012) Volume 3, pp 1-3, 2012.

[38]. Basil K Jeemon, K Shambavi, Zachariah C Alex, "A Multi-fractal Antenna for WLAN and WiMAX Application" 2013 IEEE Conference on Information and Communication Technologies (ICT 2013), pp 953-955, 2013.

[39]. V.V. Reddy, and NVSN Sarma, "Triband Circularly Polarized Koch Fractal Boundary Microstrip Antenna", IEEE Antennas and Wireless Propagation Letters Volume 13, pp 1057 - 1060, 2014.

[40]. C. Borja, C. Puente, and A. Median, “Iterative Network Model, to predict the Behavior of a Sierpinski Fractal Network”, IEEE Electronics Letters, vol. 34, pp. 1443-1445, 1998.

[41]. Carles Puente-Baliarda, Jordi Romeu, Rafael Pous, Angel Cardama, "On the Behavior of the Sierpinski Multiband Fractal Antenna" , IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, VOL. 46, NO. 4, pp 517 - 524, April 1998.

[42]. J.M Gonzalez et Al, Active zone self-similarity of fractal Sierpinski antenna verified using infra-red thermograms, Electrinics letters, Vol. 35 N 17, pp 1393-1394, 19th August 1999.

[43]. Carles Puente Baliarda, Carmen Borja Borau, Mònica Navarro Rodero, and Jordi Romeu Robert, "An Iterative Model for Fractal Antennas: Application to the Sierpinski Gasket Antenna" , IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, VOL. 48, NO. 5, pp 713-719, MAY 2000.

[44]. M. Kitlinski and R. Kieda, "Compact CPW-fed Sierpinski fractal monopole antenna", ELECTRONICS LETTERS Vol. 40 No. 22, 28th October 2004.

[45]. C. T. P. Song, Peter S. Hall, and H. Ghafouri-Shiraz, "Shorted Fractal Sierpinski Monopole Antenna", IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, VOL. 52, NO. 10, pp 2564-2570, OCTOBER 2004.

[46]. Jaume Anguera et Al, "Broadband Triple-Frequency Microstrip Patch Radiator Combining a Dual-Band Modified Sierpinski Fractal and a Monoband Antenna", IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, VOL. 54, NO. 11, pp 3367-3373, NOVEMBER 2006.

[47]. J. Vemagiri, M. Balachandran, M. Agarwal and K. Varahramyan, "Development of compact half-Sierpinski fractal antenna for RFID applications", LECTRONICS LETTERS Vol. 43 No. 22, 25th October 2007.

[48]. Kuem C. Hwang, "A Modified Sierpinski Fractal Antenna for Multiband Application", IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS, VOL. 6, pp 357-360,2007.

[49]. Wojciech J. Krzysztofik, "Modified Sierpinski Fractal Monopole for ISM-Bands Handset Applications", IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, VOL. 57, NO. 3, pp 606-615, MARCH 2009.

[50]. P.J. Soh, G.A.E. Vandenbosch, S.L. Ooi and M.R.N. Husna, "Wearable dual-band Sierpinski fractal PIFA using conductive fabric", ELECTRONICS LETTERS Vol. 47 No. 6, 17th March 2011

[51]. Federico Viani, "DUAL-BAND SIERPINSKI PRE-FRACTAL ANTENNA FOR 2.4 GHz-WLAN AND 800 MHz-LTE WIRELESS DEVICES", Progress In Electromagnetic Research C, Vol. 35, pp 63-71, 2013.

[52]. Douglas H. Werner, Randy L. Haupt, and Pingjuan L. Werner, "Fractal Antenna Engineering: The Theory and Design of Fractal Antenna Arrays", IEEE Antennas and Propagation Magazine, Vol. 41, No. 5, pp 37-59, October I999.

[53]. Ban-Leong OOI, "A Modified Contour Integral Analysis for Sierpinski Fractal Carpet Antennas with and without Electromagnetic Band Gap Ground Plane", IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, VOL. 52, NO. 5, pp 1286-1293, MAY 2004.

[54]. Rowdra Ghatak, Rabindra K. Mishra, and Dipak R. Poddar, "erturbed Sierpinski Carpet Antenna With CPW Feed for IEEE 802.11 a/b WLAN Application", IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS, VOL. 7, pp 742-744, 2008.

[55]. J. Anguera et Al, "METALLIZED FOAMS FOR ANTENNA DESIGN: APPLICATION TO FRACTAL-SHAPED SIERPINSKI-CARPET MONOPOLE", Progress In Electromagnetics Research, PIER 104, pp 239-251, 2010.

[56]. Homayoon Oraizi and Shahram Hedayati, Miniaturized UWB Monopole Microstrip Antenna Design by the Combination of Giusepe Peano and Sierpinski Carpet Fractals, IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS, VOL. 10, pp 67-70, 2011.

[57]. Balaka Biswas, D R Poddar, R. Ghatak, and Anirban Karmakar, "Modified Sierpinski Carpet Fractal Shaped Slotted UWB Monopole Antenna with Band Notch" , National Conference on Communications (NCC 2013), pp 1-5, 2013

[58]. Y. KIM and D. L. Jaggard, " The Fractal Random Array", Proc. IEEE, pp. 1278-1280, 1986.

[59]. Douglas H. Werner, Randy L. Haupt, and Pingjuan L. Werner, "Fractal Antenna Engineering: The Theory and Design of Fractal Antenna Arrays", IEEE Antennas and Propagation Magazine, Vol. 41, No. 5, pp 37-59, October I999.

[60]. Joshua S. Petko, and Douglas H. Werner, "Miniature Reconfigurable Three-Dimensional Fractal Tree Antennas", IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, VOL. 52, NO. 8, pp 1945-1956, AUGUST 2004.

[61]. Joshua S. Petko and Douglas H. Werner, "Interleaved Ultrawideband Antenna Arrays Based on Optimized Polyfractal Tree Structures", IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, VOL. 57, NO. 9, pp 2622-2631, SEPTEMBER 2009.

[62]. Javad Pourahmadazar, Changiz Ghobadi, and Javad Nourinia, "Novel Modified Pythagorean Tree Fractal Monopole Antennas for UWB Applications", IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS, VOL. 10, pp 484-48487, 2011.

[63]. Christophe Dumond, Mokhtar Khelloufi, and Levi Allam, "Experimental Study Of 2-D Electrochemically-Deposited Random Fractal Monopole Antennas", Progress In Electromagnetics Research C, Vol. 36, pp 119-130, 2013.

[64]. M. Naser-Moghadasi et Al, "UWB CPW-Fed Fractal Patch Antenna With Band-Notched Function Employing Folded T-Shaped Element", IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS, VOL. 12, pp 504-507, 2013.

[65]. C. Varadhan, J. K. Pakkathillam, M. Kanagasabai, R. Sivasamy, R. Natarajan and S. K. Palaniswamy, "Triband Antenna Structures for RFID Systems Deploying Fractal Geometry," IEEE Antennas and Wireless Propagation Letters, Vol. 12, pp. 437-440, 2013.

[66]. Guo Liu, Liang Xu, and Zhensen Wu, "ual-Band Microstrip RFID Antenna With Tree-Like Fractal Structure", IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS, VOL. 12, pp 976-978, 2013.

[67]. Dau-Chyrh Chang, Bing-Hao Zeng, , and Ji-Chyun Liu, "CPW-Fed Circular Fractal Slot Antenna Design for Dual-Band Applications", IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, VOL. 56, NO. 12, pp 3630-3636, DECEMBER 2008.

[68]. Biswajeet Mukherjee, Pragati Patel, and Jayanta Mukherjee, "Hemispherical Dielectric Resonator Antenna Based on Apollonian Gasket of Circles—A Fractal Approach", IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, VOL. 62, NO. 1, pp 40-47, JANUARY 2014.

[69]. Min Ding et Al, "Design of a CPW-fed Ultra Wideband Crown Circular Fractal Antenna ", IEEE Antennas and Propagation Society International Symposium, pp 2049 - 2052, 2006.

[70]. Raj Kumar, Yogesh B. Thakare, and M. Bhattacharya, "Novel Design ofStar Shaped Circular Fractal Antenna", International Conference on Recent Advances in Microwave Theory and Applications (MICROWAVE 2008), pp 239 - 241, 2008

[71]. Raj Kumar, and K. K. Sawant, "On the Design of Circular Fractal Antenna with U-Shape Slot in CPW-Feed ", Wireless Engineering and Technology, 1, pp 81-87, 2010.

[72]. Iman MOHAMMADSHAH, Changiz GHOBADI, and Javad NOURINIA, " A Novel Flower-Shaped Fractal Monopole Antenna with Enhancement of Bandwidth for UWB Applications", Internatıonal Journal of Natural and Engineering Sciences, pp 78-81, 2013.

[73]. V.V.Reddy and N.V.S.N.Sarma, "Compact Circularly Polarized Asymmetrical Fractal Boundary Microstrip Antenna for Wireless Applications", IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS, VOL. 13, pp 118-121, 2014.

[74]. John P. Gianvittorio and Yahya Rahmat-Samii, “Fractal Antennas: A Novel Antenna Miniaturization Technique, and Applications”, IEEE Antenna’s and Propagation Magazine, Vol. 44, No. 1, pp 20-36, February 2002.

[75]. Steven R. Best, “A Discussion on the Significance of Geometry in Determining the Resonant Behavior of Fractal and Other Non-Euclidean Wire Antennas”, IEEE Antennas and Propagation Magazine, Vol. 45, No. 3, pp 9-28,June 2003.

[76]. C. Mahatthanajatuphat, S. Saleekaw, and P. Akkaraekthalin, "A RHOMBIC PATCH MONOPOLE ANTENNA WITH MODIFIED MINKOWSKI FRACTAL GEOMETRY FOR UMTS, WLAN, AND MOBILE WIMAX APPLICATION", Progress In Electromagnetics Research, PIER 89, 57–74, 2009.

[77]. Li-Na Chen et Al,"Minkowski fractal patch antenna for size and radar cross-section reduction", 2011 IEEE CIE International Conference on RADAR, pp 1406 - 1409, 2011.

[78]. Moraes, L.B. , Barbin, S.E.,"A comparison between Minkowski and Koch fractal patch antennas", 2011 IEEE International Microwave & Optoelectronics Conference , pp 17-21, 2011.

[79]. D. K. Naji, J. S. Aziz, R. S. Fyath, "Design and Simulation of Miniaturized Minkowski Fractal Aperture-Coupled Antenna for 5.8 GHz RFID Applications ", Journal of Emerging Trends in Computing and Information Sciences, VOL. 3, NO. 7, pp 1013-1020, July 2012.

[80]. D. K. Naji, Jaber. S. Aziz, and Raad S. Fyath, “Design and Simulation of RFID Aperture Coupled Fractal Antennas", International Journal of Engineering Business Management, pp 1-14, 2012.

[81]. Suman Lata, and Varun Kuma, "Design and Simulation of Minkowski Fractal Patch Antenna on SOI Substrate for Next Generation Wireless Networks ", International Journal of Emerging Technology and Advanced Engineering , Volume 3, Issue 8, pp 92-95, August 2013.

[82]. Sayantan Dhar, Rowdra Ghatak, Bhaskar Gupta, and D.R. Poddar,"A Wideband Minkowski Fractal Dielectric Resonator Antenna", IEEE Transactions on Antennas and Propagation (Volume:61 , Issue: 6), pp 2895 - 2903, 2013.

[83]. B. H. Ahmad, H. Nornikman, M. Z. A. Abd Aziz, M. A. Othman, A. R. Othman , "Tri-band Minkowski Island Patch Antenna with Complementary Split Ring Resonator at the Ground Plane ", 13th Conference on Microwave Techniques COMITE 2013, April 17-18, Pardubice, Czech Republic, pp 46-51, 2013.

[84]. A.Singh, M.K Singh, "Design and Simulation of Miniaturized Minkowski Fractal Antennas for microwave applications", nternational Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 1, pp 5309-5311, January 2014

[85]. Ancy P V, Satya Bhushan Sukla, A K Prakash, K K Mukundan, "Multiband Fractal Antenna for wireless communication", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 3, Issue 2, pp 5753-5758, February 2014.

[86]. J. M. Gonzalez-Arbesu,S. Blanch, and J. Romeu, "THE HILBERT CURVE AS A SMALL SELF-RESONANT MONOPOLE FROM A PRACTICAL POINT OF VIEW", MICROWAVE AND OPTICAL TECHNOLOGY LETTERS / Vol. 39, No. 1, pp 45- 49, October 5 2003.

[87]. Noor Asniza Murad et Al, "Hilbert Curve Fractal Antenna for RFID Application", 2006 INTERNATIONAL RF AND MICROWAVE CONFERENCE, SEPTEMBER 12 - 14, 2006, PUTRAJAYA, MALAYSIA, pp 182-186, 2006.

[88]. I.Sanz et Al, "The Hilbert Monopole Revisited", the Fourth European Conference on Antennas and Propagation (EuCAP), pp 1-4, 2010.

[89]. Jung-Tang Huang, Jia-Hung Shiao, and Jain-Ming Wu, "A Miniaturized Hilbert Inverted-F Antenna for Wireless Sensor Network Applications", IEEE Transactions on Antennas and Propagation, (Volume:58 , Issue: 9), pp 3100 - 3103, 2010.

[90]. S. Suganthi, "Study of Compact Hilbert Curve Fractal Antennas for Implantable Medical Applications", International Journal of Emerging Technology and Advanced Engineering, Volume 2, Issue 10, pp 116-125, October 2012.

[91]. Y. S. Li, X. D. Yang, C. Y. Liu, and T. Jiang, "ANALYSIS AND INVESTIGATION OF A CANTOR SET FRACTAL UWB ANTENNA WITH A NOTCH-BAND CHARACTERISTIC", Progress In Electromagnetics Research B, Vol. 33, pp 99-114, 2011.

[92]. Gopalakrishnan Srivatsun, Sundaresan Subha Rani, Gangadaran Saisundara Krishnan, "A Self-Similar Fractal Cantor Antenna for MICS Band Wireless Applications", Wireless Engineering and Technology, Volume 2, pp 107-111, 2011.

[93]. Gopalakrishnan Srivatsun and Sundaresan Subha Rani, "Compact Multiband Planar Fractal Cantor Antenna for Wireless Applications: An Approach", International Journal of Antennas and Propagation, pp 1-6, 2012.

TNC Transactions on Networks and Communications
SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
VOLUME 2, ISSUE 5
ISSN: 2054 - 7420

Enhanced TCP Westwood Slow Start Phase

Mohanad Al-Hasanat, Kamaruzzaman Seman and Kamarudien Saadan
University Sains Islam Malaysia, Malaysia
mohanad.hasanat@gmail.com; {drkzaman, kamarudin}@usim.edu.my

ABSTRACT

Many end-to-end TCP implementations have been presented in the past decade. Despite that they used different methods to improve transport protocols over wireless networks; they mostly shared the same original TCP principles. TCP Westwood introduced a novel end-to-end bandwidth estimation mechanism. Nevertheless, it maintains the same slow start phase presented in TCP Reno. For the initial slow start phase, there is no safe slow start threshold value. In this paper, we propose to use the bandwidth estimation to calculate the initial slow start threshold value after the second round trip time. Furthermore, we introduce a faster state in which TCP increases the transmission rate once the link is underutilizing. As a result, the new proposed method shows better performance comparing to TCP Westwood, and TCP NewReno techniques.

Keywords: TCP Westwood; Congestion Control; Slow Start; Slow Start Threshold, TCP Enhancement.

1 Introduction

TCP– Transmission Control Protocol is the most transport protocol used over internet. Inefficient TCP performance in wireless networks motivated a wide spectrum in research community to enhance its congestion control mechanisms. Several TCP variants have been introduced over the past decade to support different network technologies [5, 6]. These mechanisms can be classified into three main categories: a link level solutions (e.g. I-TCP, M-TCP, etc.), end-to-end solutions (e.g. Explicit Congestion Notification (ECN), TCP Westwood, TCP Casablanca, etc.), and split connection solutions (e.g. Forward Error Correction (FEC), Automatic Repeat Request (ARQ), and Hybrid ARQ (HARQ), etc.).

TCP Westwood-TCPW presented a novel E2E bandwidth estimation mechanism by monitoring the rate of returning acknowledgments at the sender side [4]. TCPW inherent the basic TCP transmission control principles; flow control, congestion control, and error control mechanisms. The flow control tries to limit the transmission rate corresponding to the receiver's buffer capacity. Whereas, the congestion control mechanisms tries to limit the transmission rate by the link capacity. Therefore, TCP uses a congestion window (cwnd) to limit the number of segments the sender can transmit whenever a new acknowledgment received. TCPW starts the connection in slow start phase. During this phase the sender increments its cwnd exponentially until cwnd equal to a predefined value called slow start threshold (ssthresh). After that, a congestion avoidance phase is started, during which the sender increments its cwnd linearly. Anytime a packet loss event occurs, TCP sets the ssthresh value to one half the cwnd and trigger the slow start again. For the initial slow start there is no safe ssthresh value. If the ssthresh value is too small, then the sender will immediately stops the exponential increment of the cwnd. Thus, it will

DOI: 10.14738/tnc.25.601
Publication Date: 25rd October, 2014
URL: http://dx.doi.org/10.14738/tnc.25.601

take long time to reach the optimal cwnd size using the Additive Increase/Multiplicative Decrease (AIMD) linear incrementing. On the other hand, using a large ssthresh value will aggressively increase the cwnd. This as an effect causes more packet losses, unwanted retransmission events, and serious performance degradation.

In this paper, we propose to modify the TCPW bandwidth estimation in order to properly set the ssthresh value for the initial slow start phase. In addition, we present a faster start phase in which the sender can rapidly increase its cwnd size to shorten the slow start phase.

The rest of the paper is organized as follow; Section two presents a brief background study. We introduce our modifications in section three. Section four present a comparative simulation experiments to validate the proposed modifications. Then the conclusion is drawn in section five.

2 Background Study

2.1 Introduction

TCP Tahoe introduced the first congestion control mechanism in 1988 [1], including slow start, congestion avoidance, and fast retransmit. A new modification to the Tahoe's fast retransmit was presented in TCP Reno [2]. This modification used a fast recovery mechanism every time a fast retransmit procedure is triggered. Further modification to Reno is presented as TCP New-Reno [3]. TCP NewReno enhanced the fast retransmit mechanism in case of multiple packets lost from a single window.

The poor performance of TCP over wireless networks innovate a wide spectrum of research community to develop new solutions [8]. Many TCP variants have been presented to overcome this issue. One novel End-to-End bandwidth estimation method known as TCP Westwood [4] was presented. TCPW monitors the rate of the returning acknowledgments at the sender side to obtain an estimation of the link bandwidth. Then TCPW uses the estimation to set the slow start threshold value after a loss event occurs.

2.2 The Slow Start

Slow start algorithm used to gradually increase the number of packets in transit. The implementation of the slow start is accomplished through defining two variables to control the transmission; the congestion window (cwnd) and the receiver advertized window (rwnd). The cwnd is the number of packets the sender can send before receiving an acknowledgment (ACK). While the rwnd is the amount of packets the receiver can buffer. The sender limits the sending rate to the minimum of the cwnd and rwnd values.

To avoid congesting in the transmission links with large amount of data, TCP slowly probe the network capacity using slow start. Usually, TCP star transmission by cwnd=1 segment. During the slow start phase, the sender side increments cwnd by 1 segment for each ACK received. This exponential growth of cwnd ends when the cwnd exceeds the slow start threshold (ssthresh) value or when congestion observed. When packet loss event detects the value of ssthresh set to half of the cwnd size, the cwnd sets to 1 segment, and the TCP sender starts the slow start again. Figure 1 shows the cwnd growth during slow start phase.

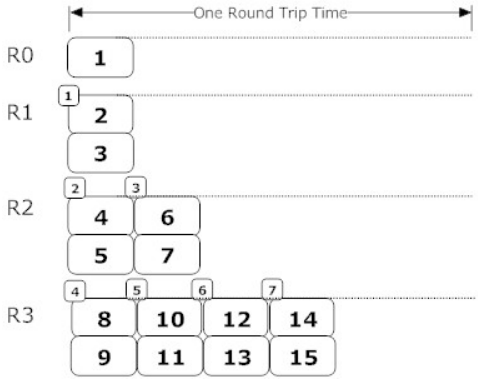


Figure 1: The Chronology of Slow Start [1]

3 The Modification

In this section we present a new method to properly set the ssthresh value in the initial slow start phase. Toward this end, we probe the link's bandwidth by counting the bytes acknowledged between two sequences ACKs at the sender side. According to the following equation:

$$ELC = \frac{Acked * SegmentSize}{t_k - t_{k-1}}$$

Where *LC* is the link capacity and *Acked* is the number of packets acknowledge within every ACK. Then we use the moving average method to update the computed ELC every ACK received according to following formula:

$$ELC = (1-\alpha)ELC_i + \alpha * ELC_{i-1}$$

Where α = 0.9.

To get the ssthresh in a form of congestion window we use the following equation:

$$ssthresh = ELC * MinRTT$$

Where RTT is round trip time (the time when a packet is sent until the ACK is received). The computed ssthresh value will provide an accurate value for initial slow start threshold according to the link capacity. As seen, this value is not a constant number; however it varies according to the connection status.

Furthermore, we proposed to use a state called the "Faster start" in which we can increase the cwnd according to a bandwidth utilization. Before sending new segment during the slow start, we check the values of the current cwnd and last round trip time. As following:

- If the last RTT is less than or equal the estimated RTT, and cwnd less than the half of ssthresh, then cwnd = cwnd + (ssthresh DIV cwnd).
- Else, cwnd =cwnd +1.

The following section shows that, the new modifications improve TCPW congestion window and the throughput.

4 Experiments and Analysis

The performance of the new modifications is assessed in this section. We use two performance metrics to evaluate the proposed modifications, throughput, and congestion window. We compare the results with TCPW and TCP NewReno. For a consistence comparison we use the same simulation scenario that had been conducted to present the original TCPW [4]. Network Simulator NS-3 is used to option the results, and gnuplot used to plot the graphs.

4.1 Simulation Setup

The topology used in this experiment is shown in Figure 2. A single source and sink connected via a gateway (PGW).

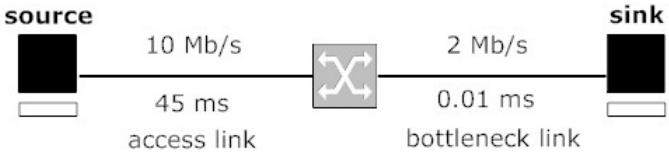


Figure 2: Simulation Topology [4].

Two links, a source-PGW link labeled access link, and PGW-sink link labeled bottleneck link. The access link bandwidth is 10Mbps with propagation delay of 45ms, where the bottleneck link bandwidth is 2Mbps with propagation delay of 0.01ms. The NS3's built-in PointToPointHelper [7] is used to represent a point to point (P2P) connection between the source-PGW and the PGW-sink. To simulate the wireless lose channel we used the RateErrorModel [7] class to generate sending errors over the bottleneck link. Errors are assumed to follow random distribution. A BulkSendApplication [7] is used to generate a single traffic along the simulation period started at the source and ended at the sink. Table1 summarizes the simulation parameters.

Table 1. Simulation parameters.

Parameter	Value
Mobility	Fixed Position
Access link bandwidth	10Mb/s
Access link Propagation Delay	45 ms
Bottleneck link bandwidth	2Mb/s
Bottleneck link Propagation Delay	0.01 ms
Error model	Uniform Error Model
Packet Error Rate (PER)	0.005
Application type	Bulk Send Application
Simulation time	5 seconds

We used 5 seconds as simulation time to focus our results on the initial slow start phase.

4.2 Simulation Results

For the first experiment, we used the same parameters listed in table 1. We compared the congestion window of the new modification referred to as Petra, TCPW, and TCP NewReno. The result is plotted in Figure 3.

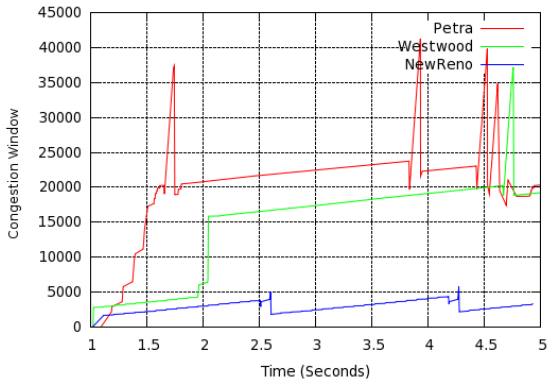


Figure 3:Congestion Window comparison.

As can be observed from this figure, a bigger cwnd values is achieved for the new modification algorithm comparing to TCPW and TCP NewReno. Moreover, we can see how fast the new modification reaches the optimal cwnd that is just about 1.75 seconds. While TCPW reached its maximum cwnd at 4.75 seconds, and TCP NewReno recorded a very small congestion window size.

Next we evaluate the total throughput achieved using the new modification, TCPW, and TCP NewReno as a function of increasing bottleneck bandwidth size. The results are plotted in Figure 4. The same network topology given above is also used.

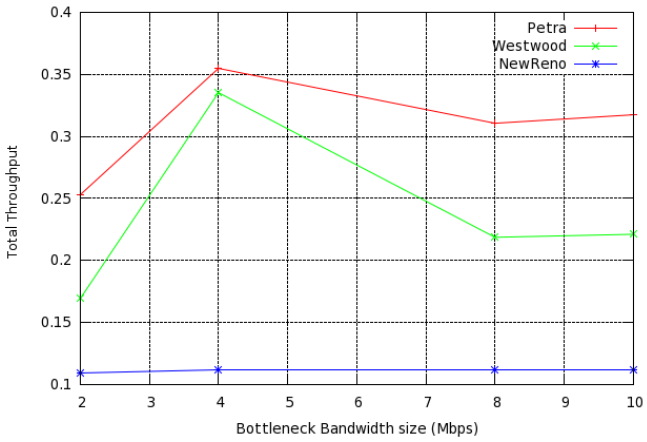


Figure 4: Throughput vs. bandwidth

It is clearly seen that a better throughput is achieved with Petra comparing to TCPW and TCP New Reno. However, weird throughput degradation is noticed after the 4Mbps bandwidth for both TCPW and Petra. One reason of this weird behavior could be due to the increasing retransmission procedures that occurred as responses to loss events. Such weird behavior did not appear for small bandwidth sizes since infrequent packet losses occurred.

The simulation experiments were extended to study the impact of various propagation delay values on the throughput. Figure 5 shows the total throughput values achieved by Petra, TCPW, and TCP New Reno over different propagation delay values started from 1ms to 60ms. The bottleneck bandwidth is set 2Mbps and the simulation period is set to 5 seconds. A BER of 0.005 is still used.

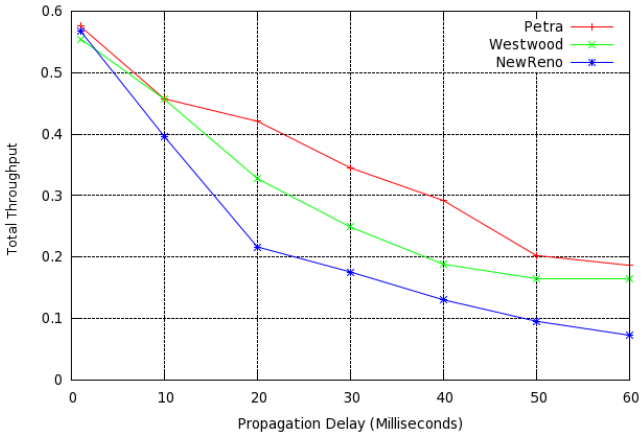


Figure 5: Throughput vs. Propagation delay

Figure 5 shows better throughput values recorded using the new modification over the other implementations. As expected, the throughput is decreased as the propagation delay is increased.

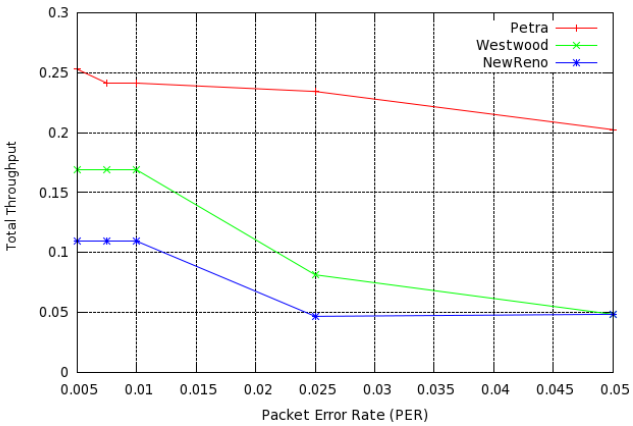


Figure 6: Throughput vs. Packets Error Rate.

In figure 6 we plotted the throughput recorded by the three implantations. We used various values of PER over the same bandwidth size and propagation delay 2Mbps and 45Ms respectively. As we can be seen, the new modification outperformed other implementations significantly for the entire range of PER.

5 Conclusion

In this paper, new modifications to TCPW slow start phase, referred to as Petra, were introduced. One modification suggested using bandwidth estimation in order to set the initial value of the slow start threshold. The other modification is represented by using a faster start to rapidly increase the congestion window size according to the link status. Simulation results in this paper slowed that, Petra improved TCP performance in terms of throughput and congestion window size.

In future work, we could further extend the evaluation process to study the impact of more performance metrics including Jitters, delay, and packet loss. Moreover, we could compare Petra to other TCP implementations, in addition to investigating the fairness and the friendless of the new modifications.

REFERENCES

[1]. Van Jacobson and M. J. Karels, *Congestion Avoidance and Control*. ACM Computer Communication, 1988. 18: p. 314-329.

[2]. V. Jacobson. *Modified TCP Congestion avoidance algorithm*, end2end-interest mailing list, April 30, 1990.

[3]. Floyd, S., et al., *The NewReno Modification to TCP's Fast Recovery Algorithm.* RFC 3782, April 2004.

[4]. S. Mascolo, C., et al., *TCP westwood: Bandwidth estimation for enhanced transport over wireless inks.* ACM SIGMOBILE, 2001. p. 287-297.

[5]. H. Xie, A., et al., *A Novel Cross Layer TCP Pacing Protocol for Multi-hop Wireless Networks*. IEEE Wireless Communications and Networking Conference, 2013.

[6]. C. Hu, X., et al., *WiTracer: A Novel Solution to Improve TCP Performance overWireless Network.* IEEE, 2013.

[7]. The ns-3 Network Simulator Doxygen Documentation. http://www.nsam.org/doxygen/group_tcp.html, December 2013.

[8]. M. Al-Hasanat, K., et al.. Enhanced TCP Westwood Congestion Control Mechanism over Wireless Networks. In International Conference on Advanced Technology & Sciences. 12-15 August, 2014. Antalya, T

TNC Transactions on Networks and Communications
SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
VOLUME 2, ISSUE 5
ISSN: 2054 - 7420

Enhancing the competence of enterprise network using contemporary networking paradigms

Aditya Ahuja, Kamal Dewan, Nikita Gupta and Meenakshi Sood
ECE Department, Jaypee University of Information Technology, Waknaghat Solan, India;
adityaahuja2005@gmail.com; kdewan9495@gmail.com; nikita92gupta@yahoo.com;
meenakshi.sood@juit.ac.in

ABSTRACT

Networking has traversed from days where networks were considered a background component of businesses to today's electronic age, where networks are an imperative resource, and directly determine revenue generation for an organization. In today's dynamic arena of networking, the crafting of networks has escalated from using just an elemental set of features, to consolidating modernistic technologies and services, in an effort to come up with state of the art networks which can meet the aim of connectivity, security, scalability, simplicity of operation, and flexible accommodation of new trends and technologies. The roots of this paper lie in the Enterprise network, and while adhering to the need of the hour in corporate sector, we propose architecture for Enterprise network, using avant-garde technologies such as Frame Relay, Port Security, Access Control Lists (Firewalling), VoIP, VPN, Ether Channel, Redistribution of Routing Protocols and ISP Redundancy. The network architecture has been designed on Cisco's network simulation software: Cisco Packet Tracer. The principle behind the proposed network architecture can be applied in designing the networks of a host of other campuses.

Keywords: Frame Relay, Port Security, ACL (Access Control List), Ether Channels, VoIP (Voice over Internet Protocol), VPN (Virtual Private Network).

1 Introduction

The designing of networks has evolved and matured from merely applying a basic set of techniques, to incorporating multiple technologies and services, in an effort to support the vastly disparate end to end communication requirements.

In the past, network designers had a very limited number of options in terms of hardware devices, protocols and media, and thus network designing was relatively easier, with very little scope of mistake, but with limited efficiency and flexibility. Whereas, today's networks are based on complex environments, which are an amalgamation of multiple protocols, media and interconnections to networks outside any single organization's dominion of control, thus giving rise to computationally efficient networks, which can scale, and flexibly accommodate upgrades without an entire revamp of the design. Networks are broadly classified as *LAN (Local Area Network)*, and *WAN (Wide Area Network)*. LANs, which persist over a relatively shorter distance are designed to allow personal computers to share resources, which can include hardware (e.g., a printer), software (e.g., an application program), or data.

DOI: 10.14738/tnc.25.565
Publication Date: 31rd October, 2014
URL: http://dx.doi.org/10.14738/tnc.25.565

A WAN, is an amalgamation of LANs which are spread over large geographical areas. WAN provides transmission of data over large distances that may encompass a country, a continent, or even the whole world [1].

This paper resides on the Enterprise Network, which is a building or a group of buildings, all connected into one central network that consists of many LANs. In enterprises today, more business is conducted electronically and deals are closed rapidly. Thus, in today's digital age, company operations have undergone a sea-change, and 24*7 connectivity has never been more imperative than it is today. Therefore, it is apt to say that the enterprise network has matured from an inert business element to a very active and visible asset that today's organizations rely on to support their day-to-day functions. It is seen as a critical resource, which directly supports revenue generation. When the network is going to interface with the internet, its security is also an important aspect, thus, today's networks must be open and pervasive, yet remain secure and controlled [2]. Moreover, the demand for mobile computing has increased in today's business environment, thus the networks must also be accessible remotely [3]. Therefore, new enterprise network designs are needed, as heirloom solutions and techniques cannot meet the new requirements, nor reduce the costs and streamline the operations [2].

This paper focuses on designing a state of the art enterprise network, by effectively blending a plethora of new technologies such as VoIP, VPN, Frame Relay, Port Security, Ether Channel, Access Control Lists (Firewalling), Redistribution of Routing Protocols and ISP Redundancy. The proposed architecture has been implemented and tested on Cisco's Network Simulation Program: *"Cisco Packet Tracer"*.

The rest of the paper is organized as follows: section 2 presents a brief insight into the technologies incorporated in the proposed design. Section 3 demonstrates the proposed network architecture, as designed on Cisco Packet Tracer. Section 4 presents the results and discussions along with demonstration of some of the results.

2 Technologies Incorporated

2.1 LAN Framework Design

The present scenario in the corporate world demands networks that can support high-speed business solutions such as voice, video, wireless, and mission-critical data applications. This calls for highly dependable and security-centred network designs. Using the principles of hierarchical design fulfils the requirements for building such efficient networks. The hierarchical design rests upon a building block approach. It consists of various independent distribution blocks which are attached to a high-speed routed core network layer [4].

The hierarchical design incorporates three layers, the *access* layer, *distribution* layer and the *core* layer. The Access Layer is where the end users are allowed in to the network. The Distribution Layer will contain switches and routers capable of VLAN switching and allow defining departmental workgroups and multicast domains. The Core Layer is capable of switching packets as fast as possible.

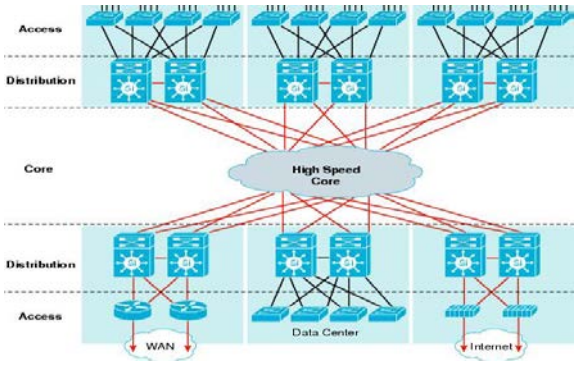


Figure 1: Hierarchical Model of LAN design

This design approach splits the functions of the network into various building blocks to provide reliability, flexibility, scalability, and fault isolation, thus making it a cluster of multiple smaller, more manageable hierarchical blocks. With the advancement in technology, the focus was highly on network convergence. The basic hierarchical model was modified to adapt itself with the active development of new services such as Voice over Internet Protocol (VoIP), Virtual Private Network (VPN), Frame Relay, Ether Channel, Port Security, etc. Myriads of such features, which the authors have also incorporated in their proposed design, have been discussed below.

Virtual Private Network (VPN) is a network that normally uses internet to establish connection between two branches of an enterprise. The security and protection of the shared information is maintained using special tunneling protocols and complex encryption procedures, hence the new connection so established is a dedicated point-to point connection. Besides costing significantly lesser as compared to privately owned or leased services, the service also provides data integrity, thus making it a private network virtually.

Voice over Internet Protocol (VoIP) is a rapidly surfacing technology that merges the voice and data networks for voice communication, using the all-pervasive IP-based networks to deploy VoIP client devices such as IP phones, mobile VoIP-enabled handheld devices, and VoIP gateways. Conventional telecommunications systems, even Private Branch Exchanges (PBX) are rapidly being replaced by IP-based systems (IP PBX) and thus, Voice over IP is extensively entering the enterprise networks [8].

Frame Relay is a packet-switched technology, which allows multiple sites of an organization, located within a few kilometers, to connect. All the locations plug into the frame relay “cloud”, which is usually a conglomeration of dozens or hundreds of Frame-Relay switches and routers. Unlike Ethernet switches, which make decisions based on MAC addresses, Frame Relay switches make decisions based on Data Link Connection Interfaces (DLCI). For communication to occur between locations, virtual circuits (VC) must be created, which is a one-way path through the Frame-Relay cloud. Virtual circuits are identified with DLCIs. Frame-relay circuits can either be permanent, or switched. A Switched Virtual Circuit (SVC) is created only when traffic needs to be sent, and is torn down when communication is complete. The authors have incorporated Permanent Virtual Circuit (PVC) in their designed network, which is always kept active, and is the most common virtual circuit. PVCs are software defined, so they can be created, altered or dismantled in a matter of hours. Networks working on the principle of frame relay have the higher speed and lower delay qualities of circuit switching without the need for dedicated full-time devices and circuits and wasted time slots when no data is being transmitted [9]. Also, Frame relay

networks are considered private because each customer's individual traffic is separated into a predetermined path, the PVC. Every PVC has an associated Committed Information Rate (CIR) that defines the amount of bandwidth a customer is provided on the shared network facility. However, customers have the ability to transmit data on their PVC at rates up to the full port speed [10].

Ether Channel technology, built upon standards-based 802.3 full-duplex Fast Ethernet, has emerged as a reliable, high-speed solution for the campus network backbone. It allows grouping of multiple ports into a single logical transmission path between a switch and a router, server, or another switch [11]. Besides making fair distribution of traffic between the channels, the technology also provides redundancy in the event of link failure. If a link is cut in an Ether Channel, traffic is rerouted to one of the other links in less than a few milliseconds [5].

Port security is a mechanism available on switches to restrict access to the devices that can connect via a particular port of the switch. Port security activated on a switch port only allows machines with a MAC address belonging to the range configured on it to connect to the switch's network. The MAC address of a frame arriving on the switch port is compared with the MAC addresses configured in its allowed list. The packet is allowed to pass through if its MAC address matches. If the MAC address is not a member of the configured list, the port either drops the packet or shuts itself down for a considerable amount of time [6].

Firewall is either hardware or software based security mechanism in a network. A hardware based firewall is a dedicated device with its own operating system on a specialized platform, whereas a software-based firewall or *Access Control List (ACL)* is an additional program loaded on a network device like a router to inspect data or network traffic. As a check point gateway, it analyses the IP packets and decides whether to allow through or not, based on the preconfigured rules. An ACL also determines which information or services to be accessed from outside as well as from inside and by whom. The authors have implemented firewall in the form of ACL in their proposed architecture.

2.2 Routing Protocols Used

Routing is an integral part of IP network design because it is the mechanism that provides accessibility for the applications. The communicating devices need to agree on a common set of rules to share the information. Such set of rules are known as protocols.

Routing Information Protocol (RIP) is a distance-vector routing protocol. This protocol has the feature of sending out the complete routing table to all active interfaces regularly and whenever the network topology changes. When a router receives a routing update that includes changes to an existing stored entry, it updates its routing table incorporating the new route. The protocol uses *hop count* to measure the distance between the source and destination network. It prevents routing loops from continuing indefinitely by limiting the number of hops allowed in a path to *15*. RIP version 1 uses classful routing, which means that the same subnet mask is used by all devices in the network. In the designed network, the authors have used RIP version 2 which utilizes classless routing. *Open Shortest Path First (OSPF)* is an open link-state routing protocol. It sends out link-state advertisements (LSAs) on attached interfaces, to all other routers within the same hierarchical area. After the link-state information has been accumulated, they use the Shortest Path First algorithm to calculate the shortest path to each node of the network. *Enhanced Interior Gateway Routing Protocol (EIGRP)* incorporates the capabilities of link-

state protocols into distance vector protocols. Under this protocol, no periodic updates are carried out. Upon the detection of a route change, partial updates are sent out. Partial updates are sent only to those routers that need the information. Fast convergence is another factor that distinguishes the protocol from other routing protocols. A router running EIGRP stores all its neighbors' routing tables so that it can quickly adapt to alternate routes [7]. EIGRP is apt for very large networks, having a maximum hop count of 255. As in the network architecture proposed in the paper, redistribution is the requirement of a network running different routing protocols. While running a single routing protocol throughout the entire IP internetwork is desirable, multi-protocol routing may be an outcome of company mergers, multiple departments being managed by multiple network administrators, or multi-vendor environments.

3 Proposed Architecture

The proposed network architecture has been depicted in Fig. 2.The different departments of the organisation, namely IT Operations, Marketing, Human Resource, and Finance, are located in Block 1 and Block 2. Each of the departments is under different VLANs. Each department has been provided an IP phone. Block 3 houses the organisation’s server room, which is the area where all the servers supporting the organisation’s network have been placed. Separate servers have been provided for the organisation’s website, data storage and for DNS service. Block 4 is the area where the Research and Development (R&D) department is located. To increase bandwidth and provide link redundancy, ether channels have been implemented in this area. For employees’ recreation, Cafeteria 1 and Cafeteria 2 have been designed, and Wi-Fi access has been provided. In response to a situation of the link to the primary ISP breaking down, the network incorporates a link from a secondary ISP, which has been administratively turned down. In case the primary ISP fails, the secondary ISP can be pushed into service. Another feature of the network is the application of Frame Relay, to interconnect all the offices of the organisation in the same city or state, those located within smaller distances. Site to site VPN over the internet has been created, to interconnect the organisation’s offices located at larger distances.

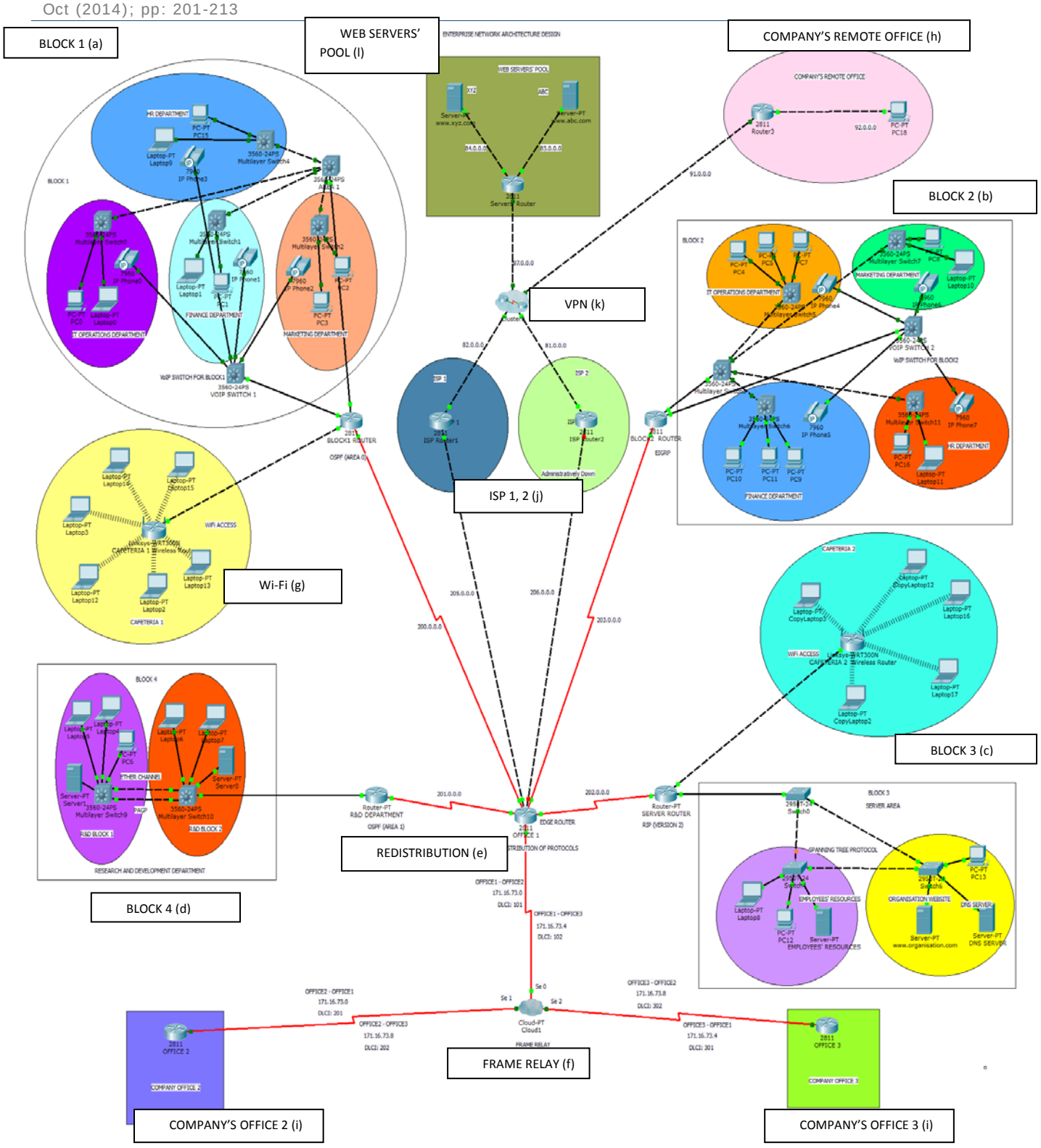


Figure 2: Proposed Network Architecture

4 Results and Discussions

The proposed network is *DHCP* (Dynamic Host Configuration Protocol) enabled. The programming modules prepared for different parts of the network are shown in Fig. 3-9.

Block 1, as shown in Fig. 2(a) is the area where one leg of all the different departments, namely, IT Operations, Marketing, Human Resource, and Finance, are placed. A separate switch for VoIP facility has been used, to which the IP phones for each department are connected. Configuration for VoIP has been depicted in Fig. 3. *Cafeteria 1* is located next to Block 1, and Wi-Fi access has been provided here, as shown in Fig. 2(g) for conducting informal meetings and for employee recreation. OSPF (Open Shortest Path First) routing protocol has been implemented in these areas. Block 1 and Cafeteria 1 are under OSPF area 0.

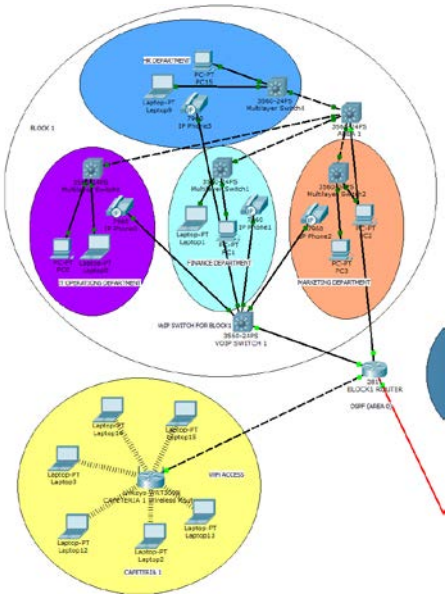


Figure 2(a): Block 1

Block 2, shown in Fig. 2(b) is the area where the second leg of the organisation's departments, along with VoIP Phones, has been placed. *Cafeteria 2* has been designed next to Block 2. *EIGRP (Enhanced Interior Gateway Routing Protocol)* has been implemented in these areas.

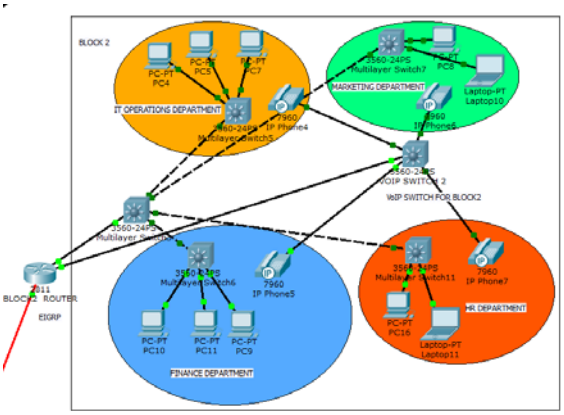


Figure 2(b): Block 2

The servers supporting the organisation's network are present in *Block 3*, which is the *Server Area*, running on *RIPv2* (Routing Information Protocol version 2), depicted by Fig. 2(c). One server is the Employees' resources server, second is the DNS server, and third is the organisation's website server. To provide undeterred access to the servers even in situations where any one of the paths goes down, a redundant path has been provided which has been automatically blocked by *STP (Spanning Tree Protocol)*, to avoid loop formation.

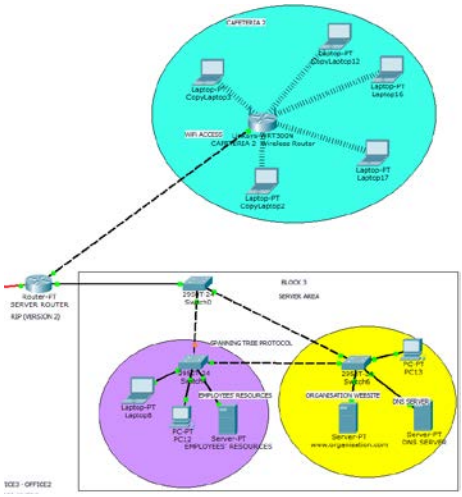


Figure 2(c): Block 3

Block 4, shown in Fig. 2(d) houses the *R&D* (*Research and Development) Department* of the organisation. This area has further been divided into R&D Block 1 and R&D Block 2. In order to shield the organisation's R&D activities from leaking out, ACL has been deployed on the R&D department router, as shown in Fig. 4. Block 4 is thus secured, as neither a user from within the R&D department can reach any other part of the organisation's network or the internet, nor vice versa. Moreover, *Port Security* has been implemented on the switches in the R&D department, as depicted by Fig. 5. The MAC addresses of the verified users' computers have been configured on the ports of the switches, so that an alien user cannot connect to the switch. Also, *Ether Channels* have been deployed between the two switches, i.e. two individual Ethernet links have been bundled into a single logical link. If a segment within an Ether Channel fails, traffic previously carried over the failed link switches to the second segment within the Ether Channel. This has been done to guarantee greater bandwidth as well as uninterrupted communication between the members of the two blocks. The Ether Channels have been configured using the *Port Aggregation Control Protocol (PAgP)*. The configuration has been shown by Fig. 6. Block 4 is running on OSPF and is a part of Area 1.

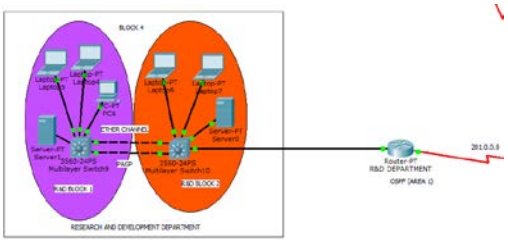


Figure 2(d): Block 4

As the design is an amalgamation of multiple protocols, *redistribution* is necessary, and has been implemented on the edge router, shown in Fig. 2(e), to enable inter-protocol communication between the three protocols, i.e. OSPF, EIGRP and RIPv2.

```
Router#conf t
Router(config)#telephony-service //activating telephony services on the router
Router(config-telephony)#no auto-reg-ephone
Router(config-telephony)#ip source-address 10.0.0.1 port 2000
Router(config-telephony)#max-ephones 10
Router(config-telephony)#max-dn 100
Router(config-telephony)#create cnf-files
Router(config-telephony)#exit
Router(config)#ephone-dn 1 //creating a telephone number for an IP Phone
Router(config-ephone-dn)#number 1000
Router(config-ephone-dn)#exit
Router(config)#ephone-dn 1
Router(config-ephone-dn)#exit
Router(config)#ephone 1
Router(config-ephone)#mac-address 0009.7C08.C930 //registering the telephone number to a particular IP Phone
Router(config-ephone)#exit
Router(config)#dial-peer voice 1 voip //enabling inter-network calls on IP Phones
Router(config-dial-peer)#destination-pattern ....
Router(config-dial-peer)#session target ipv4:200.0.0.1
Router(config-dial-peer)#exit
Router(config)#router ospf 1 //adding the networks of IP Phones to their respective routing configurations
Router(config-router)#network 10.0.0.0 0.0.0.255 area 0
Router(config-router)#network 20.0.0.0 0.0.0.255 area 0
Router(config-router)#network 30.0.0.0 0.0.0.255 area 0
Router(config-router)#exit
```

Figure 3: Configuration of VoIP on Block 1 Router

```
Router#conf t
Enter configuration commands, one per line.  End with CNTL/Z.
Router(config)#access-list 10 deny any //defining an access controllist to deny every network
Router(config)#int f0/0
Router(config-if)#ip acc
Router(config-if)#ip access-group 10 out //applying the ACL to the interface connecting R&D department
Router(config-if)#exit
```

Figure 4: Configuration of Access Control List

```
Switch#conf t
Switch(config-if)#switchport mode access
Switch(config-if)#switchport port-security
Switch(config-if)#switchport port-security maximum 1 //defines the maximum number of devices which can be attached
Switch(config-if)#switchport port-security mac-address sticky //adds currently attached device's MAC address to list
Switch(config-if)#switchport port-security violation shutdown
Switch(config-if)#exit
```

Figure 5: Configuration of Port Security in R&D Department

```
Switch(config)#interface  port-channel 3 //configuring the Ether Channels
Switch(config-if)#switchport  trunk encapsulation  dot1q
Switch(config-if)#switchport  mode trunk
Switch(config-if)#exit
Switch(config)#interface  range f0/3-4
Switch(config-if-range)#channel-protocol  pagp
Switch(config-if-range)#channel-group  3 mode desirable
Switch(config-if-range)#exit
//The corresponding channel is completed by configuring the same on the opposite switch
```

Figure 6: Configuration of Ether Channels in R&D Department

```
Router#conf t //Redistribution of EIGRP with RIP and OSPF
Enter configuration commands, one per line.  End with CNTL/Z.
Router(config)#router eigrp 2
Router(config-router)#redistribute rip metric 1000 10 255 255 100
Router(config-router)#redistribute ospf 3 metric 1000 10 255 255 100
Router(config-router)#exit
Router(config)#
//Redistribution of OSPF with EIGRP and RIP, and RIP with OSPF and EIGRP is also carried out
```

Figure 7: Redistribution of routing protocols on edge router

It has been assumed that offices 1, 2 and 3 of the organisation are situated in the city A, with other offices in far-away states and countries. To connect the three offices in the same city, the authors have used *Frame Relay*, shown in Fig. 2(f), as this would lead to cost savings as well as provide high bandwidth according to the *CIR* (Committed Information Rate). *PVCs* have thus been created between each of the three offices, therefore guaranteeing secure and fast transmission of data between the offices. The cloud represents the Frame Relay network and the organisation's offices 2 and 3 have each been depicted by a router, for representational purpose. The configuration has been illustrated in Fig. 8(a)-(b).

```
Router#conf t
Router(config)#int s1/7
Router(config-if)#no ip address
Router(config-if)#encapsulation frame-relay
Router(config-if)#int s1/7.1 point-to-point  //creating sub-interface
Router(config-subif)#frame-relay interface-dlci 101 //providing DLCI to the sub interface
Router(config-subif)#ip address 171.16.73.1 255.255.255.252 //assigning IP address to the sub-interface
Router(config-subif)#int s1/7.2 point-to-point
Router(config-subif)#frame-relay interface-dlci 102
Router(config-subif)#ip address 171.16.73.5 255.255.255.252
Router(config-subif)#exit
```

Figure 8(a): Configuration for Frame Relay

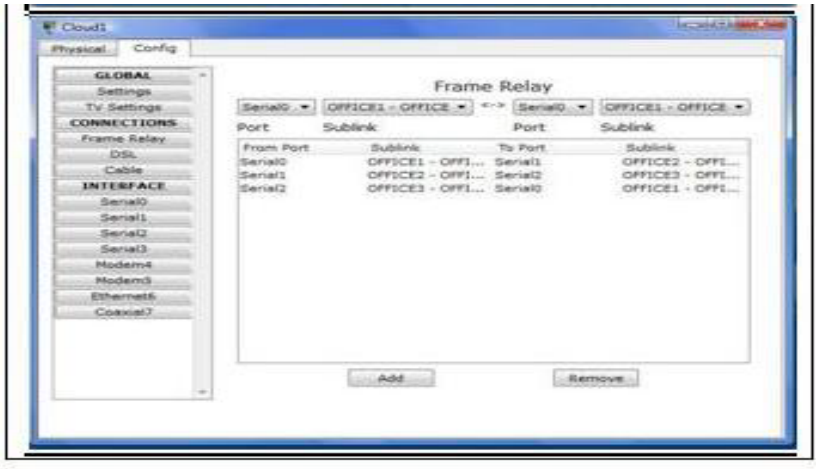


Figure 8(b): Configuration for Frame Relay

ISP (Internet Service Provider) *1* and *ISP 2*, as shown in Fig. 2(j), represent the internal networks of the two Internet Service Providers for the organisation. The link to ISP 2 has been kept in shut-down mode

administratively, and would be activated only in case of the link to the primary ISP shutting down. The network incorporates a connection to the (redundant) secondary ISP, to guarantee uninterrupted internet access for the organisation.

```
Router#configure terminal //configuring VPN tunnel from remote office (Router 3)
Router(config)#interface FastEthernet0/0
Router(config-if)#ip address 91.0.0.2 255.0.0.0
Router(config-if)#exit
Router(config)#router rip
Router(config-router)#version 2
Router(config-router)#network 92.0.0.0
Router(config-router)#network 91.0.0.0
Router(config-router)#exit
Router(config)#crypto isakmp policy 10
Router(config-isakmp)#authentication pre-share
Router(config-isakmp)#encryption aes 256
Router(config-isakmp)#group 2
Router(config-isakmp)#lifetime 86400
Router(config-isakmp)#exit
Router(config)#crypto isakmp key toor address 91.0.0.1
Router(config)#crypto ipsec transform-set TSET esp-aes esp-sha-hmac
Router(config)#access-list 101 permit ip 92.0.0.0 0.255.255.255 82.0.0.0
0.255.255.255
Router(config)#crypto map CMAP 10 ipsec-isakmp
Router(config-crypto-map)#set peer 91.0.0.1
Router(config-crypto-map)#match address 101
Router(config-crypto-map)#set transform-set TSET
Router(config-crypto-map)#exit
Router(config)#int f0/0
Router(config-if)#crypto map CMAP
*Jan 3 07:16:26.785: %CRYPTO-6-ISAKMP_ON_OFF: ISAKMP is ON
//Configuration of VPN tunnel is also carried out on the Block 1 Router as well as
all the intermediate routers.
```

Figure 9: Configuration for VPN

The *Web Servers' Pool*, shown in Fig. 2(l), represents the servers of two websites, namely, *www.xyz.com* and *www.abc.com*. This area is running on RIPv2.

Company's Remote Office, as shown in Fig. 2(h), represents one of the offices of the organisation, in a different region or country. A *site-to-site VPN*, shown in Fig. 2(k), has been established between the remote office and the office under consideration, configuration as shown by Fig. 9. A secure *IPSec tunnel* has been set up starting from the remote office's router to the Block 1 router, for representational purposes. Traffic between the two offices is transmitted over the internet at best effort. Therefore, remote offices of the organisation are connected to each other as though they are a part of the same network.

4.1 Demonstration

The network, when simulated on Cisco Packet Tracer, produced the following results:

(a). On trying to ping a member of the R&D Department from another part of the network, the result "Destination Host Unreachable" gets displayed. The result has been illustrated in Fig. 10(a).

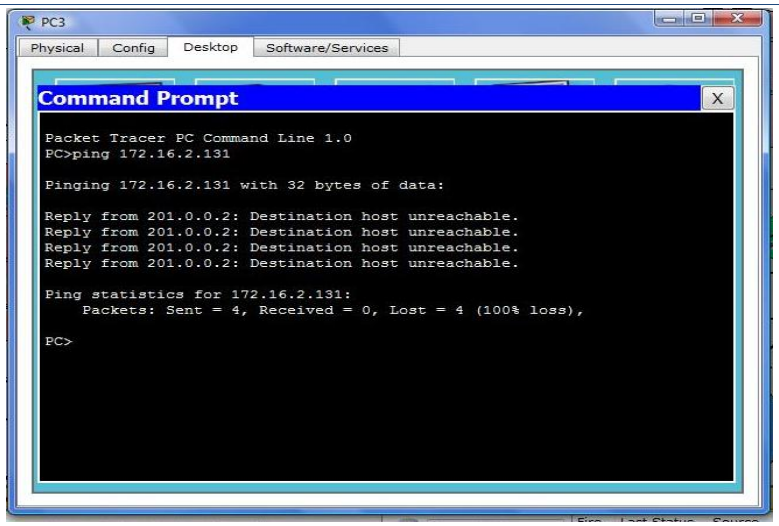


Figure 10(a): Unreachable host of R&D Department

(b). When a member of the R&D Department tries to ping a member of say, Block 1, "Request Timed out" gets displayed, as depicted by Fig. 10(b).

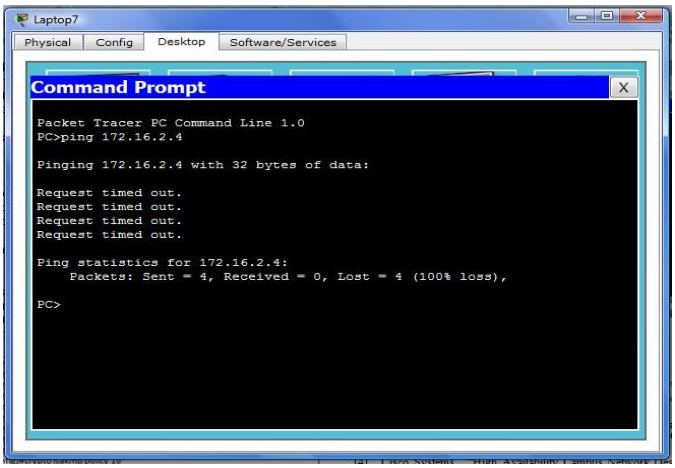


Figure 10(b): Host of another Block inaccessible from R&D Block

Thus, it can be concluded that the R&D Department has been well shielded from the outside network by the ACL.

Thus, winding up in a nutshell, the fact that nowadays in an organization, peer-to-peer communication is a more important part of Local Area Networks than client-server communication, led to the incorporation of Voice over Internet Protocol in the proposed design. Moreover, VoIP is advantageous as it leads to cost savings and poses no geographical boundaries. In case of the connection to the primary ISP breaking down, a redundant connection to a secondary ISP has been provisioned to guarantee 24*7 internet connectivity. The R&D department in the designed network has been supplied with Ether Channel technology to guarantee the best availability of resources and its functionality under the most severe circumstances. The isolation of R&D department from the other parts of the organization and the exterior world could be possible due to implementation of software firewall in the form of Access Control List. The sensitive company data on the servers of R&D department has been

secured by incorporation of Port Security in the area, which will prevent anyone to disconnect the presently connected computers, connect any unauthorized device and hack the data. Also, established as a network of privately owned equipment, Frame Relay is able to connect multiple offices of the organization in the same city. Site to site VPN is able to utilize the flexibility and ubiquity of the Internet in connecting remote offices of the organisation, for the holistic working and growth of the entire enterprise.

REFERENCES

[1]. Andrew S. Tanenbaum, 2003, Computer Networks, Prentice Hall PTR

[2]. Campus LAN Design Guide: Design Considerations for the High-Performance LAN, Juniper Networks, Inc., 2009

[3]. Martin W. Murhammer, Kok-Keong Lee, Payam Motallebi, Paolo Borghi, Karl Wozabal, IP Network Design Guide, IBM Corporation, 1999

[4]. Cisco Systems, "High Availability Campus Network Design--Routed Access Layer using EIGRP or OSPF", 2007

[5]. Qutaiba Ali, Salah Alabady, and Yehya Qasim, "Applying reliability solutions to a cooperative network," International Arab Journal of e-Technology, Vol. 1, No. 2, pp. 9-17, June 2009

[6]. Saadat Malik, Network Security Principles and Practices: Expert solutions for securing network infrastructures and VPNs, Cisco Press, 2003

[7]. Cisco Systems, Internetworking Technology Handbook, 2012

[8]. T-Systems, White paper- "Voice over Internet Protocol (VoIP)"

[9]. Hewlett Packard – "Frame Relay Networks", Digital Technical Journal, Vol. 5, No. 1, Winter 1993

[10]. Sprint, White paper- "Frame Relay vs. IP VPNs"

[11]. Cisco Systems, "Scaling Networks Companion Guide", 2014